

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *NAIVE BAYES*
PADA PILKADA KABUPATEN SITUBONDO TAHUN 2024**

SKRIPSI



Oleh:

MA'RUF UBAIDILLAH

2021503113

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY
SITUBONDO**

2025

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *NAIVE BAYES*
PADA PILKADA KABUPATEN SITUBONDO TAHUN 2024**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan Program
Sarjana (S-1) pada Program Studi Teknologi Informasi Fakultas Sains dan
Teknologi Universitas Ibrahimy



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY
SITUBONDO
2025**

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : **Ma'ruf Ubaidillah**
NPM/NIRM : 2021503113
Program Studi : S-1 Teknologi Informasi
Fakultas : Dakultas Sains dan Teknologi

Menyatakan dengan sebenarnya, bahwa tugas akhir/skripsi ini secara keseluruhan adalah hasil penelitian atau karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sebagai sumber referensi dan disebutkan dalam daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa tugas akhir/skripsi ini hasil plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Situbondo, 29 Agustus 2025

Saya yang menyatakan,



Ma'ruf Ubaidillah

PERSETUJUAN PEMBIMBING

Nama : **Ma'ruf Ubaidillah**
NPM/NIRM : 2021503113
Program Studi : **Analisis Sentimen Menggunakan Algoritma *Naive Bayes* Pada PILKADA Kabupaten Situbondo Tahun 2024**

Telah disetujui oleh:

Pembimbing 1



Achmad Baijuri, M.Kom.
NIDN: 715078902

Pembimbing 2



Nur Azise, M.Kom.
NIDN: 730108802

PENGESAHAN


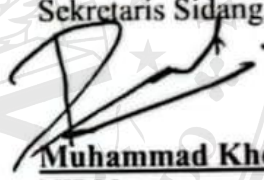
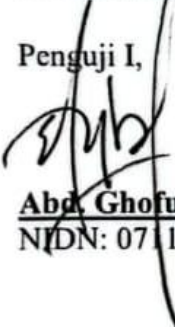
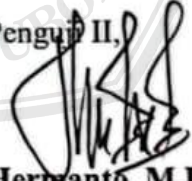
SKRIPSI

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA NAIVE BAYES
PADA PILKADA KABUPATEN SITUBONDO TAHUN 2024**


Ma'ruf Ubaidillah
2021503113

telah dipertahankan di depan dewan penguji Sidang/Munaqasyah Skripsi pada hari Sabtu,
Tanggal 23 agustus 2025 sebagai salah satu syarat memperoleh gelar Sarjana (S.Kom)
pada Fakultas Sains dan Teknologi Universitas Ibrahimiy.

Tim Penguji,

Ketua Sidang,  <u>Dr. Ach. Khumaidi, M.P.</u> NIDN: 0722049001	Sekretaris Sidang,  <u>Muhammad Khoirroni, S.Kom</u> NIDN: -
Penguji I,  <u>Abd. Ghofur, M.Kom.</u> NIDN: 0711088303	Penguji II,  <u>Hermanto, M.Kom.</u> NIDN: 708087807

Mengetahui.

Dekan,

Abd. Ghofur, M.Kom.
NIDN: 0711088303

MOTTO

“Berhenti Karena Keterbatasan Atau Berkarya Dengan Segala Kekreatifitasan”



PERSEMBAHAN

1. Superhero yang akan selalu menjadi panutan hidup, Ayahanda tercinta, terimakasih selalu berjuang untuk kehidupan penulis, beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, Namun beliau mampu mendidik penulis, memotivasi, memberikan dukungan hingga penulis mampu menyelesaikan studinya sampai sarjana.
2. Pintu surgaku, Ibunda tercinta, yang tidak henti-hentinya memberikan kasih sayang dengan penuh cinta dan selalu memberikan motivasi serta do'a hingga penulis mampu menyelesaikan studinya sampai sarjana.
3. Tak kalah pentingnya saya ucapkan terima kasih kepada diri sendiri yang mampu menyelesaikan studi ini meski banyak rintangan yang perlu dihadapi, dan terus percaya bahwa segala persoalan pasti ada solusinya. Tetap optimis dan berdiri sendiri atas keyakinan diri yang diyakini.
4. Terimakasih kepada kakak serta keluarga untuk doa, support dan pengertiannya selama ini.
5. Terimakasih kepada kakak-kakak tingkat yang sudah berbaik hati meluangkan waktu dan bimbingannya.
6. Terimakasih kepada teman-teman di Kampus Universitas Ibrahimi Kota yang selalu memberikan support serta motivasi, terkhusus kepada Gus Ahmad Gunz dan Syekh Ali Yasin meski selalu "absen".
7. Terakhir tak lupa saya ucapkan terimakasih kepada semua yang telah bersedia membaca skripsi ini.

KATA PENGANTAR

Segala puji syukur peneliti sampaikan kepada Allah SWT, karena atas Rahmat dan Hidayah-Nya, perencanaan, pelaksanaan dan penyelesaian tugas akhir/skripsi dengan judul “Analisis Sentimen Menggunakan Algoritma *Naive Bayes* Pada PILKADA Kabupaten Situbondo Tahun 2024” sebagai salah satu syarat penyelesaian program diploma/sarjana dapat terselesaikan dengan baik dan lancar.

Kesuksesan ini dapat peneliti peroleh karena dukungan beberapa pihak. Peneliti menyampaikan terima kasih kepada :

1. KHR. Ach. Azaim Ibrahimi, S. Sy, M. HI Selaku Pengasuh Pondok Pesantren Salafiyah Syafi'iyah Sukorejo Situbondo.
2. KH. Ach. Fadlail, SH, M.H Selaku Rektor Universitas Ibrahimi.
3. Abd. Ghofur, M. Kom Selaku Dekan Fakultas Sains dan Teknologi Universitas Ibrahimi.
4. Dr. Ach. Khumaidi, M.P Selaku Wakil Dekan 1 Fakultas Sains dan Teknologi Universitas Ibrahimi.
5. Abd. Wafi, M.P Selaku Wakil Dekan 2 Fakultas Sains dan Teknologi Universitas Ibrahimi.
6. Ahmad Lutfi, M. Kom Selaku Wakil Dekan 3 Fakultas Sains dan Teknologi Universitas Ibrahimi.
7. Firman Santoso, M.Kom Selaku Ketua Program Studi Teknologi Informasi
8. Achmad Baijuri, M.Kom. dan Nur Azise, M.Kom. selaku pembimbing I dan II, yang telah memberikan arahan dan bimbingan
9. Seluruh Dosen Fakultas Sains dan Teknologi yang telah memberikan kami ilmu sehingga sampai pada masa Skripsi ini.

Semoga semua amal baik yang telah diberikan oleh Bapak/Ibu kepada peneliti mendapat balasan yang sebaik mungkin dari Allah SWT, Amin.

Situbondo, 17 Agustus 2025

Penyusun

ABSTRAK

Pemilihan kepala daerah (PILKADA) Kabupaten Situbondo tahun 2024 menjadi salah satu momen penting dalam dinamika demokrasi daerah. Media sosial, khususnya *TikTok*, digunakan secara luas oleh masyarakat untuk menyampaikan opini, baik positif, netral, maupun negatif terhadap kandidat. Namun, banyaknya komentar tanpa penyaringan menimbulkan beragam persepsi yang sulit dipetakan. Kondisi ini mendorong penelitian untuk menganalisis sentimen masyarakat melalui pendekatan *data mining*, sehingga opini publik dapat dikaji secara lebih objektif dan bermanfaat bagi strategi komunikasi politik.

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimental. Data penelitian diperoleh melalui teknik *web scraping* dari 494 komentar *TikTok* yang membahas PILKADA Situbondo. Data yang terkumpul diproses melalui tahapan *text preprocessing* seperti *cleansing*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya, data diklasifikasikan menggunakan algoritma *Multinomial Naive Bayes* dengan pembagian 80% data latih dan 20% data uji. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* serta metrik akurasi, *presisi*, *recall*, dan *f1-score*.

Hasil penelitian menunjukkan bahwa model klasifikasi berbasis *Naive Bayes* mampu mengklasifikasikan sentimen dengan akurasi sebesar 70,59%. Kinerja model lebih baik dalam mengidentifikasi sentimen positif dibandingkan netral, yang masih sulit diklasifikasikan secara tepat. Temuan ini membuktikan bahwa metode *Naive Bayes* cukup efektif dalam pemetaan opini publik berbasis media sosial, meskipun masih terdapat ruang pengembangan melalui optimalisasi algoritma atau penambahan variasi dataset. Temuan ini dapat dimanfaatkan oleh kandidat, tim kampanye, maupun penyelenggara pemilu dalam merancang strategi komunikasi politik yang lebih tepat sasaran.

Kata Kunci: Analisis Sentimen, *Naive Bayes*, *TikTok*, PILKADA Situbondo, *Data Mining*

ABSTRACT

The 2024 Regional Head Election (PILKADA) in Situbondo Regency represents a crucial democratic momentum at the local level. Social media, particularly TikTok, has become a major platform where citizens express their opinions, either positive, neutral, or negative, toward political candidates. However, the massive number of unfiltered comments creates diverse perceptions that are difficult to categorize. This study aims to conduct sentiment analysis on TikTok comments to map public opinion more objectively and provide valuable insights for political communication strategies.

This research employed a quantitative approach with an experimental design. A total of 494 TikTok comments related to the Situbondo PILKADA were collected through web scraping techniques. The data underwent text preprocessing stages, including cleansing, case folding, tokenizing, stopword removal, and stemming. The classification was carried out using the Multinomial Naïve Bayes algorithm with an 80% training set and 20% testing set. Model performance was evaluated using a confusion matrix along with accuracy, precision, recall, and f1-score metrics.

The results show that the Naïve Bayes model achieved a classification accuracy of 70.59%. The algorithm performed better in recognizing positive sentiments compared to neutral ones, which remained more challenging to identify accurately. These findings indicate that Naïve Bayes is a relatively effective method for sentiment mapping on social media, although improvements are still possible through algorithm optimization and larger datasets. The outcomes of this study can be utilized by candidates, campaign teams, and election organizers to design more targeted political communication strategies.

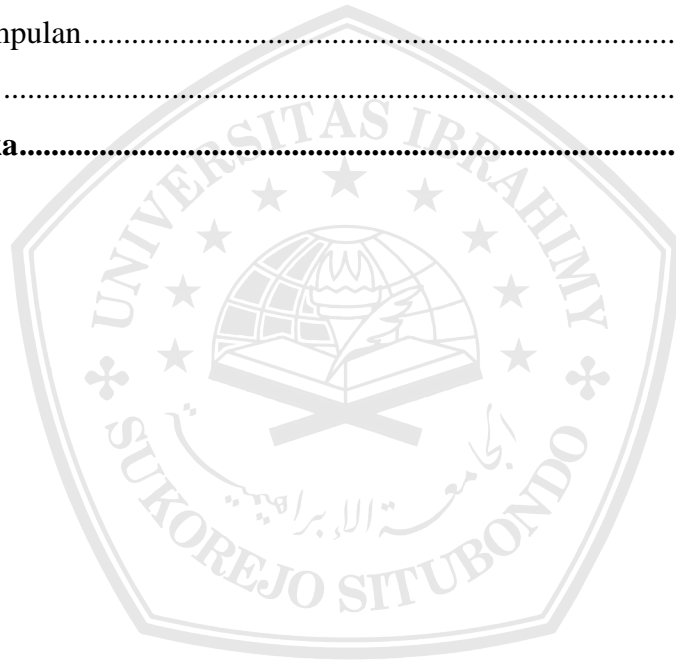
Keywords: *Sentiment Analysis, Naïve Bayes, TikTok, Situbondo Election, Data Mining*

DAFTAR ISI

HALAMAN JUDUL	ii
PERNYATAAN KEASLIAN TULISAN	iii
PERSETUJUAN PEMBIMBING	iv
PENGESAHAN	v
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR RUMUS	xvi
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Batasan Masalah	4
1.5 Tujuan Penelitian.....	4
1.6 Manfaat Penelitian.....	4
1.7 Metode Penelitian.....	5
1.7.1 Jenis Penelitian.....	5
1.7.2 Metode Pengumpulan Data	6
1.8 Sistematika Pembahasan	7
BAB II	9
TINJAUAN PUSTAKA	9
2.1 Penelitian Terdahulu.....	9

2.1.1	Analisis Sentimen Hasil Putusan MK Terkait Sengketa PILPRES 2024 Pada Media Sosial X Dengan Metode <i>Naive Bayes</i>	9
2.1.2	Analisis Sentimen Pada Ulasan Produk Pakaian Zirdigo Dengan Menggunakan Algoritma <i>Naive Bayes</i>	10
2.1.3	Analisis Sentimen Pengguna Aplikasi <i>Shopee</i> Pada Situs <i>Google Play</i> Menggunakan <i>Naive Bayes Classifier</i>	12
2.2	Landasan Teori	13
2.2.1	<i>Data Mining</i>	13
2.2.2	Klasifikasi	14
2.2.3	Analisis Sentimen	14
2.2.4	Algoritma <i>Naive Bayes</i>	15
2.2.5	<i>Confusion Matrix</i>	15
2.2.6	<i>Python</i>	17
2.3	Pemodelan	17
2.4	Perangkat Lunak yang Digunakan	19
2.4.1	<i>Microsoft Excel</i>	19
2.4.2	<i>TikTok</i>	19
2.4.3	<i>Rapid Miner</i>	19
2.4.4	<i>Data Miner</i>	20
2.4.5	<i>Visual Studio Code</i>	21
BAB III	22
ANALISIS PERANCANGAN	22
3.1	Gambaran Umum Objek Penelitian	22
3.1.1	<i>Scrapping Data TikTok</i>	22
3.2	Alur Proses	26
3.2.1	Identifikasi dan Analisis proses Bisnis	26
3.2.2	Analisis Proses Bisnis	27
3.2.3	Identifikasi dan Analisis Kebutuhan	31
3.2.4	Identifikasi dan Analisis Alternatif Solusi	32
3.2.5	Desain Proses	33
BAB IV	37

HASIL DAN PEMBAHASAN	37
4.1 Kebutuhan Sistem.....	37
4.2 <i>Preprocessing</i> Data	37
4.3 Wordcloud	44
4.4 Pelabelan	46
4.5 Split Data Training dan Data Testing.....	50
4.6 Metode <i>Naive Bayes Clasifier</i>	52
BAB V.....	60
PENUTUP.....	60
5.1 Kesimpulan.....	60
5.2 Saran	61
Daftar Pustaka.....	62



DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	16
Tabel 2. 2 Simbol <i>Flowchart</i>	18
Tabel 3. 1 Identifikasi Alternatif Solusi	32
Tabel 3. 2 Analisis Kelayakan Alternatif Solusi	33
Tabel 4. 1 Proses <i>Cleaning</i>	38
Tabel 4. 2 Case Folding	40
Tabel 4. 3 <i>Tokenizing</i>	41
Tabel 4. 4 <i>Remove Stopword</i>	42
Tabel 4. 5 <i>Stemming Data</i>	43



DAFTAR GAMBAR

Gambar 3. 1 Buat Resep Data Miner	23
Gambar 3. 2 Pilih Jenis Resep.....	24
Gambar 3. 3 Menetapkan Baris.....	24
Gambar 3. 4 Menetapkan Kolom dan Menentukan Nama Jenis.....	25
Gambar 3. 5 Hasil Scrapping	25
Gambar 3. 6 Hasil Berbentuk File excel	26
Gambar 3. 7 Tahapan Algoritma <i>Naive Bayes</i>	30
Gambar 3. 8 Preprocessing Data.....	34
Gambar 3. 9 Alur Labeling Data.....	34
Gambar 3. 10 Alur Metode <i>Naive Bayes</i>	35
Gambar 3. 11 Arsitektur Sistem.....	36
Gambar 4. 1 Wordcloud Sebelum Preprocessing	45
Gambar 4. 2 Frekuensi Kata Sebelum Preprocessing	45
Gambar 4. 3 Wordcloud Setelah Preprocessing.....	46
Gambar 4. 4 Frekuensi Kata Setelah Preprocessing	46
Gambar 4. 5 Hasil Labeling Data.....	47
Gambar 4. 6 Diagram Batang Klasifikasi Sentimen	47
Gambar 4. 7 <i>Wordcloud</i> Sentimen Positif	48
Gambar 4. 8 Frekuensi Kata Sentimen Positif.....	48
Gambar 4. 9 <i>Wordcloud</i> Sentimen Negatif.....	49
Gambar 4. 10 Frekuensi Kata Sentimen Negatif	49
Gambar 4. 11 <i>Wordcloud</i> Sentimen Positif	50
Gambar 4. 12 Frekuensi Kata Sentimen Netral	50
Gambar 4. 13 Diagram Batang Split Data.	51
Gambar 4. 14 Hasil Prediksi Pada Data Sentimen.....	53
Gambar 4. 15 Confusion Matrix	54
Gambar 4. 16 Classification Report.....	55

DAFTAR RUMUS

Rumus 2.1 Naive Bayes	15
Rumus 2.2 Akurasi	17
Rumus 2.2 Precision	17
Rumus 2.2 Recall	17
Rumus 2.2 F1-Score	17
Rumus 3.1 Probabilitas Prior	28
Rumus 3.2 Conditional Probability	29
Rumus 3.3 Variabel Maximum A Posteriori	29



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada tahun 1998 gerakan reformasi secara eksplisit mengamanatkan pemilihan kepala daerah di Indonesia, atau biasa disingkat PILKADA. Dengan adanya keharusan bagi masyarakat untuk berpartisipasi aktif dalam memilih pemimpin daerah, pemilihan kepala daerah telah muncul sebagai momentum demokrasi yang penting dalam sejarah negara-bangsa Indonesia[1]. PILKADA di Kabupaten Situbondo tahun 2024 adalah salah satu masa krusial ketika masyarakat akan memilih pemimpin yang dianggap dapat melanjutkan kepemimpinan daerah atau membawa perubahan yang signifikan. Media sosial telah muncul sebagai tempat utama di era digital saat ini bagi individu untuk menyuarakan pemikiran *netizen*, terlibat dalam wacana politik, dan mengkomunikasikan tujuan *netizen* dengan mengacu pada politisi saingan.

Salah satu *platform* media sosial yang digunakan oleh banyak orang di Indonesia adalah *TikTok*. *TikTok* menjadi *platform* sebagai sarana interaksi sosial, hiburan, dan ekspresi diri sehingga penggunaannya terus meningkat. *TikTok* juga menjadi salah satu situs jejaring sosial yang paling banyak digunakan di antara pengguna internet di seluruh dunia [2]. Selain memberikan hiburan, *platform* ini berfungsi sebagai saluran untuk penyebaran informasi[3], termasuk yang berkaitan dengan masalah politik. Karena banyaknya video, komentar, dan tagar *TikTok* yang berkaitan dengan pemilu menimbulkan berbagai wacana publik yang hidup dan mengungkapkan opini tentang kandidat atau topik tertentu. Sehingga dibutuhkan

analisis sentimen terhadap data *TikTok* untuk memeriksa teks digital, mengidentifikasi sentimen atau emosi yang ada di dalamnya, dan memastikan apakah pendapat *netizen* positif, netral, atau negatif[4].

Teknologi *data mining* memiliki peran penting dalam proses pengumpulan data dengan jumlah yang besar. *Naive Bayes* menjadi metode dalam penelitian analisis sentimen *TikTok* pada PILKADA di Kabupaten Situbondo. Metode *Naive Bayes Classifier* adalah teknik *data mining* yang mudah dengan kinerja yang cepat dalam melatih data, kemudahan penggunaan, dan memiliki efisiensi yang tinggi[5]. *Naive Bayes* adalah jenis pengklasifikasi jaringan *Bayesian* yang paling sederhana. Setiap node fitur dalam *Naive Bayes* memiliki node kelas sebagai induknya, tetapi tidak ada node fitur lain yang memiliki orang tua. Meskipun demikian, banyak penelitian dalam pembelajaran terarah telah menunjukkan bahwa algoritma *Naive Bayes* yang paling dasar sekalipun dapat menyaingi pengklasifikasi yang canggih seperti C4.5 dan berada di 10 besar metode *data mining* secara global.[6] seperti penelitian yang dilakukan oleh Dafa Setyo Nugroho, dkk. dengan judul Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan *Tweet* Menggunakan Algoritma *Naive Bayes Classifier*, menganalisis komentar *negative* dan komentar *positive*. Dari penelitian tersebut diperoleh nilai sebesar 84.13% pada akurasi, sebesar 77.36% pada *precision*, dan nilai *recall* sebesar 91.51% dari hasil klasifikasi ulasan Twitter. Dalam konteks PILKADA di Kabupaten Situbondo, penggunaan *Naive Bayes* untuk menganalisis sentimen di *TikTok* diharapkan dapat membantu mengidentifikasi tren opini publik, memberikan wawasan kepada para

kandidat mengenai berbagai persepsi *netizen* dengan memetakan komentar yang positif, netral, dan negatif.

Memberikan gambaran tentang bagaimana *netizen* memandang PILKADA merupakan tujuan dari penelitian ini. Dalam rangka memahami opini publik dengan lebih baik dan meningkatkan taktik komunikasi politik para kandidat, staf kampanye, dan penyelenggara PILKADA dapat mengambil manfaat besar dari temuan analisis sentimen. Oleh karena itu, penelitian ini sangat relevan dengan PILKADA di Kabupaten Situbondo. Selain membantu pengelolaan data opini publik, hal ini juga memungkinkan para analis untuk memetakan topik yang paling banyak dibicarakan, memahami sepenuhnya dinamika opini publik, dan memprediksi pergeseran sentimen pada masa selanjutnya yang dapat memengaruhi hasil pemilu[7].

1.2 Identifikasi Masalah

Berdasarkan paparan latar belakang diatas, ditemukan identifikasi masalah yaitu komentar netizen di *platform TikTok* tentang PILKADA di Kabupaten Situbondo, pada komentar tersebut terdapat komentar positif, negatif, dan netral. Hal tersebut menyebabkan tidak ada *filtering* sehingga menimbulkan berbagai opini publik baik tentang calon kandidat maupun program yang akan dilaksanakan kandidat terpilih.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan indentifikasi masalah di atas, rumusan masalah dalam penelitian ini yang akan dibahas yaitu “Bagaimana membuat program aplikasi untuk mengetahui performa algoritma *Naive Bayes* dalam

mengklasifikasikan komentar *netizen TikTok* terkait PILKADA Kabupaten Situbondo tahun 2024?”

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini dikarenakan cakupan teori *Data Mining* cukup luas untuk dibahas, sehingga memuat beberapa batasan sebagai berikut:

- a. Dataset didapatkan melalui beberapa postingan di *platform TikTok* yang memuat tentang PILKADA di Kabupaten Situbondo tahun 2024 dan memiliki respon *netizen* yang cukup banyak.
- b. Dataset yang telah diperoleh diklasifikasi menggunakan algoritma *Naive Bayes* untuk menentukan pendapat atau komentar positif, netral, serta negatif.

1.5 Tujuan Penelitian

Membuat program aplikasi untuk mengimplementasikan algoritma *Naive Bayes* dalam analisis sentimen pada PILKADA di Kabupaten Situbondo tahun 2024 di *platform TikTok* menjadi tujuan penelitian ini. Dengan memahami opini publik lebih baik dan dapat meningkatkan taktik komunikasi politik para kandidat, staf kampanye, dan penyelenggara PILKADA dapat mengambil manfaat besar dari temuan analisis sentimen.

1.6 Manfaat Penelitian

Analisis sentimen menggunakan algoritma *Naive Bayes* memiliki beberapa manfaat:

- a. Dapat membantu dalam meningkatkan dan memahami opini publik sebagai taktik komunikasi politik melalui penerapan teknologi *Data Mining*.
- b. Memberikan informasi terkait persepsi masyarakat terhadap PILKADA di Kabupaten Situbondo pada masa selanjutnya.

1.7 Metode Penelitian

1.7.1 Jenis Penelitian

a. Kuantitatif

Deskriptif kuantitatif merupakan jenis penelitian yang digunakan dalam penelitian ini. Penelitian deskriptif kuantitatif adalah teknik yang menggambarkan, menjelaskan, atau meringkas berbagai pengaturan, situasi, kejadian, atau variabel penelitian berdasarkan kejadian aktual yang dapat diamati, difoto, diwawancarai, dan didokumentasikan[8].

b. Eksperimental

Eksperimental merupakan salah satu jenis penelitian yang berusaha untuk menunjukkan bagaimana sebuah perlakuan yang mempengaruhi terhadap efeknya. Menurut Arikunto, eksperimen adalah suatu metode di mana peneliti dengan sengaja menciptakan suatu kejadian atau keadaan dan kemudian meneliti hasilnya, dengan kata lain penelitian dilakukan dengan menciptakan suatu kejadian atau keadaan. Penelitian eksperimen adalah metode untuk menentukan hubungan sebab akibat antara dua variabel dengan cara menemukan hubungan sebab-akibat antara dua elemen melalui eksperimen[9]. Eksperimen yang dilakukan dalam penelitian ini adalah

menggunakan bahasa pemrograman Python untuk menganalisis penggunaan algoritma Naive Bayes pada dataset yang telah didapatkan.

1.7.2 Metode Pengumpulan Data

Pada metode penelitian ini terdapat beberapa metode pengumpulan data yang digunakan diantaranya :

a. *Web Scraping*

Web Scraping bertujuan untuk mengumpulkan data secara otomatis dari situs web untuk diperiksa. Informasi diekstraksi dari beberapa situs web dan diubah menjadi format yang dapat digunakan untuk analisis data. Hal ini menghilangkan kebutuhan pengumpulan data secara manual dan memberikan kita akses ke informasi yang lebih terkini dan akurat. Di sini, “manual” mengacu pada mengakses web, membaca, dan menarik kesimpulan. *Web Scraping* sangat penting untuk analisis data otomatis karena memungkinkan data dikumpulkan dari beberapa sumber untuk analisis tambahan[10].

b. Studi Literatur

Studi literatur adalah suatu metode penelitian dengan cara perbandingan, mengulas, dan perangkuman informasi dari sejumlah penelitian yang relevan, baik dari buku-buku atau publikasi penelitian sebelumnya yang mendukung sumber penelitian[11]. Dengan menggunakan teknik literatur ini, peneliti memanfaatkan data sekunder dari beberapa *e-journal* yaitu *Google Scholar*, buku-buku dan jurnal-jurnal penelitian terdahulu yang mendukung dan membantu dalam pembuatan laporan.

1.8 Sistematika Pembahasan

Sistematika pembahasan dalam karya tulis ini dibagi menjadi lima bab, yang memiliki suatu tujuan tertentu dalam setiap bab:

BAB I : PENDAHULUAN

Dalam pendahuluan memuat tentang latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika pembahasan.

BAB II : TINJAUAN PUSTAKA

Pada bab dua dipaparkan tentang tinjauan pustaka, hal-hal yang terkait serta referensi penunjang yang sesuai dengan judul penelitian yang diangkat, juga memuat landasan teori dan perangkat lunak yang akan digunakan.

BAB III : METODELOGI PENELITIAN

Bab ini membahas tentang bagaimana peneliti mendapatkan data yang digunakan dalam penelitian ini dan bagaimana cara mendapatkan, serta proses sebelum melakukan analisa lebih lanjut menggunakan algoritma *Naive Bayes*.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini adalah hasil dari analisa yang telah diputuskan sebelumnya. Berisi tentang beberapa penjelasan maupun perhitungan dari data yang telah diolah sehingga menghasilkan sebuah prediksi berdasarkan kemungkinan yang telah didapatkan sebelumnya dan juga tingkat akurasi dari prediksi yang telah dihasilkan.

BAB V : PENUTUP

Kesimpulan yang merangkum temuan penelitian utama, seperti hasil akurasi model, efektivitas teknik yang digunakan, atau pola signifikan yang

teridentifikasi dalam data, semua telah disesuaikan untuk memenuhi tujuan studi. Bagian ini juga memberikan rekomendasi pada penelitian selanjutnya, seperti penggunaan algoritma baru, pengintegrasian data yang lebih beragam, pengujian dalam berbagai kondisi, atau penerapan sistem dalam praktik untuk penilaian tambahan.



BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

2.1.1 Analisis Sentimen Hasil Putusan MK Terkait Sengketa PILPRES 2024 Pada Media Sosial X Dengan Metode *Naïve Bayes*

Penelitian dilakukan oleh Afifah Nurul Izzati dari Program Studi Teknik Informatika, Universitas Kahuripan Kediri. Dipublikasikan pada 01 Mei 2024 di Jurnal Teknologi Informasi Indonesia (JTII).

Pada penelitian tersebut topik yang dibahas adalah tentang pemilihan presiden dan wakil presiden Indonesia, atau PILPRES, adalah tahap penting dalam proses demokrasi ketika warga negara dapat memilih pemimpin negara. Pemilihan presiden Indonesia pada tanggal 14 Februari 2024 menimbulkan kegembiraan di kalangan masyarakat. Namun, hasil pemilu tersebut menimbulkan kontroversi dan reaksi yang tidak menyenangkan, yang berujung pada gugatan yang menggugat hasil tersebut. Dalam hal ini, Mahkamah Konstitusi (MK) memainkan peran penting dalam menyelesaikan sengketa pemilu, dan legitimasi serta kepercayaan publik terhadap pemerintah dan sistem hukum dipengaruhi oleh keputusannya.

Algoritma *Naïve Bayes Classifier* diterapkan dalam penelitian tersebut sebagai bagian dari strategi *data mining*. Menurut peneliti pendekatan ini dipilih karena *Naïve Bayes* merupakan algoritma klasifikasi dengan kualitas independen yang dapat membantu dalam kategorisasi data dan bersifat lugas, cepat, dan mudah digunakan. Langkah pertama dalam proses ini adalah mengumpulkan data ulasan dari media sosial, yang kemudian diproses terlebih dahulu agar siap untuk dilabeli

ke dalam kategori sentimen (positif, netral, dan negatif). Untuk menggunakan algoritma *Naïve Bayes*, yang berusaha menilai dan mengkategorikan sentimen publik terkait putusan MK dalam menyelesaikan sengketa PILPRES 2024, data kemudian dipisahkan menjadi data latih dan data uji.

Penelitian menghasilkan bahwa dari 145 titik data yang dikumpulkan dan diperiksa, persepsi publik terhadap putusan Mahkamah Konstitusi dalam isu PILPRES 2024 secara umum kurang baik. Dari 102 hasil tes yang disertakan, 31 di antaranya bersikap positif, 15 netral, dan 56 bersikap negatif. Tingkat tidak puasan masyarakat yang signifikan terhadap putusan Mahkamah Konstitusi tercermin dalam kesimpulan ini, yang dapat berdampak pada kepercayaan publik terhadap pemerintah dan sistem peradilan. Dalam kerangka demokrasi Indonesia, penelitian ini menekankan betapa pentingnya memahami opini publik sebagai reaksi terhadap putusan pengadilan yang signifikan[12].

2.1.2 Analisis Sentimen Pada Ulasan Produk Pakaian Zirdigo Dengan Menggunakan Algoritma *Naive Bayes*

Penelitian dilakukan oleh Cep Lukman Rohmat, Rizky Tito Aprilianto, Fathurrohman, dan Iin dari Teknik Informatika, STMIK IKMI Cirebon. Dipublikasikan pada 1 Februari 2024 di Jurnal JATI (Jurnal Mahasiswa Teknik Informatika).

Penelitian tersebut berfokus pada pertumbuhan *e-commerce* dan teknologi informasi yang telah mengubah cara orang berinteraksi dan mendapatkan informasi, sehingga analisis sentimen menjadi sangat penting untuk memahami evaluasi pelanggan yang mempengaruhi keputusan untuk membeli. Dengan

semakin banyaknya evaluasi produk di platform internet, terutama untuk produk pakaian Zirdigo, sangat penting untuk melihat sikap positif dan negatif dalam ulasan-ulasan ini untuk menilai kualitas produk dan layanan. Tujuan dari penelitian ini adalah untuk lebih memahami opini konsumen melalui analisis yang tepat dengan menggunakan algoritma *Naive Bayes*, yang dapat memberikan informasi penting kepada para pedagang tentang cara meningkatkan kualitas barang dan pada akhirnya mempengaruhi kebahagiaan pelanggan.

Melalui analisis sentimen, penelitian ini menawarkan pengetahuan yang lebih dalam tentang opini dan preferensi konsumen terhadap produk pakaian Zirdigo, yang dapat membantu para peritel dalam menentukan kualitas barang dan jasa. Studi ini menawarkan informasi yang akurat dan dapat dipercaya dengan mengklasifikasikan ulasan ke dalam perasaan positif atau negatif menggunakan algoritma *Naive Bayes*. Hal ini memungkinkan pedagang untuk membuat penilaian yang lebih baik dalam meningkatkan produk dan layanan berdasarkan umpan balik dari pelanggan. Selain itu, temuan dari penelitian ini berpotensi untuk digunakan sebagai peta jalan untuk pengembangan teknik analisis sentimen yang lebih canggih di masa depan, yang akan menjadi kemajuan besar dalam bidang analisis data dan pemrosesan bahasa alami.

Menurut temuan penelitian, analisis sentimen terhadap evaluasi produk pakaian Zirdigo menggunakan algoritma *Naive Bayes* menghasilkan akurasi sebesar 81,00%, presisi 100%, dan *recall* 80%. Hal ini menunjukkan bahwa model tersebut dapat membedakan antara evaluasi yang baik dan negatif, yang mengindikasikan bahwa ulasan yang dikumpulkan dari situs *e-commerce Shopee*

telah dikategorikan secara akurat dan tepat. Hasil ini menunjukkan seberapa baik algoritma *Naive Bayes* bekerja untuk analisis sentimen dan menekankan betapa pentingnya umpan balik pelanggan untuk membantu pedagang dalam memahami bagaimana pelanggan melihat produk, yang akan mengarah pada peningkatan kualitas dan layanan di masa depan[13].

2.1.3 Analisis Sentimen Pengguna Aplikasi *Shopee* Pada Situs *Google Play* Menggunakan *Naive Bayes Classifier*

Penelitian dilakukan oleh Andriani Nurian, Muhammad Samsul Ma'arif, Indira Nur Amalia, dan Chaerur Rozikin dari Universitas Singaperbangsa Karawang, Jawa Barat. Dipublikasikan pada 1 Januari 2024 JITET (Jurnal Informatika dan Teknik Elektro Terapan).

Untuk mengidentifikasi ulasan pengguna dari *Google Play Store*, penelitian ini menggunakan pendekatan *Naive Bayes Classifier* untuk menganalisis sentimen pada aplikasi *e-commerce Shopee*. Dengan mengumpulkan 3000 ulasan pengguna dan mengkategorikannya ke dalam kelompok negatif, positif, dan netral, penelitian ini berusaha untuk memahami bagaimana pengguna mengevaluasi aplikasi *Shopee*. Pengumpulan data, *preprocessing*, transformasi teks, dan klasifikasi adalah beberapa langkah dalam proses analisis. Temuan klasifikasi kemudian dievaluasi untuk memberikan wawasan berharga tentang kebahagiaan pengguna dan opini aplikasi.

Pengetahuan yang lebih baik tentang sikap pengguna tentang aplikasi *Shopee* adalah salah satu keuntungan dari penelitian ini, dan ini dapat menjadi dasar untuk meningkatkan kegunaan dan tingkat layanan aplikasi. Klasifikasi akurasi

tinggi dari evaluasi pelanggan dalam penelitian ini ke dalam kategori positif, negatif, dan netral menggunakan algoritma *Naïve Bayes Classifier* memberikan informasi penting kepada pengembang *Shopee* tentang area yang perlu diperbaiki. Selain itu, dengan menangani komentar pengguna, menunjukkan area masalah, dan mengembangkan rencana perbaikan yang lebih efisien, temuan studi ini dapat membantu bisnis meningkatkan kebahagiaan dan loyalitas pengguna aplikasi.

Hasil dari penelitian ini menunjukkan hasil algoritma dalam mengkategorikan evaluasi pengguna, dengan analisis sentimen *Naïve Bayes Classifier* terhadap pengguna aplikasi *Shopee* yang mencatat akurasi 87,58%, presisi 91,20%, dan *recall* 90,21%. Informasi tersebut dibagi menjadi 2.525 ulasan buruk, 276 ulasan baik, dan 199 ulasan netral dari 3000 ulasan yang berhasil dikumpulkan. Selain itu, hasil visualisasi *wordcloud* memberikan ringkasan dari istilah-istilah yang sering digunakan dalam ulasan, termasuk “*Shopee*”, “harga”, dan “potong”, yang menunjukkan perhatian utama konsumen saat memberikan komentar. Selain menerangi emosi dan pengalaman pengguna, data ini menawarkan wawasan yang dapat digunakan oleh bisnis untuk meningkatkan fitur dan penawaran pada aplikasi *Shopee*[14].

2.2 Landasan Teori

2.2.1 *Data Mining*

Praktik mengidentifikasi pola dan informasi berharga dalam volume data yang sangat besar dikenal sebagai *data mining*. Statistik data, ekstraksi data, analisis data, dan pengumpulan data semuanya termasuk dalam *data mining*. Nama lain dari *data mining* adalah sebagai *information harvesting*, *Knowledge extraction*,

Knowledge discovery, data/pattern analysis. Metode lain yang masuk akal untuk menemukan informasi yang relevan juga disebut *data mining*. Setelah ditemukan data informasi dan tren dapat digunakan sebagai instrumen tambahan untuk pengambilan keputusan di perusahaan yang akan dikembangkan. Banyak masalah terkait bisnis yang terlalu sulit untuk diselesaikan dapat diselesaikan dengan menggunakan teknik *data mining*. Selain itu, *data mining* dapat digunakan untuk memprediksi tren masa depan, yang memungkinkan para pebisnis untuk membuat keputusan yang dinamis, proaktif, dan efektif[15].

2.2.2 Klasifikasi

Salah satu metode utama dalam bidang *machine learning* adalah klasifikasi, yang mencoba mengelompokkan data ke dalam kelompok atau kelas yang telah ditentukan sebelumnya sesuai dengan sifat-sifatnya. Menemukan pola atau korelasi yang mendasari data dan menggunakan pengetahuan ini untuk meramalkan data baru dengan kategori yang tidak diketahui adalah tujuan utama klasifikasi[16].

2.2.3 Analisis Sentimen

Analisis sentimen merupakan proses pengumpulan informasi tentang pikiran, sikap, dan emosi dari teks baik itu dari makalah, ulasan produk, tweet, dan jenis komunikasi lainnya. Meskipun sering digunakan secara bergantian, sentimen dan pendapat pada dasarnya berbeda. Sentimen adalah emosi atau perasaan yang diungkapkan seseorang tentang sesuatu, baik netral, negatif, maupun positif. Di sisi lain, pendapat adalah pandangan baik positif maupun negatif yang diungkapkan seseorang tentang sesuatu[17].

2.2.4 Algoritma *Naive Bayes*

Algoritma *Naive Bayes* menggunakan teknik statistik dan probabilitas untuk mengkategorikan variabel tertentu. Dengan memeriksa frekuensi setiap klasifikasi dalam data pelatihan, *Naive Bayes* menggunakan konsep matematika yang disebut teori probabilitas untuk menentukan kemungkinan tertinggi dari suatu klasifikasi[18].

Berikut ini merupakan rumus yang digunakan dalam perhitungan *Naive Bayes* untuk menentukan kemungkinan atau probabilitas penentuan klasifikasi :

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

- a. X merupakan data sampel dengan kelas yang tidak diketahui
- b. H ialah hipotesa yaitu X adalah data dengan kelas (label) C.
- c. P(H) adalah peluang dari hipotesa H
- d. P(X) adalah peluang data sampel yang diamati
- e. P(X|H) merupakan peluang dari data sampel X, bila hipotesa benar (valid).

2.2.5 *Confusion Matrix*

Salah satu metode untuk menilai model klasifikasi dan menentukan apakah sebuah objek benar atau salah adalah metodologi *Confusion Matrix*. Untuk menghitung *Confusion Matrix* merupakan rumus *Confusion Matrix multiclass* (terdiri lebih dari 2 dataset) yang jelas berbeda dengan rumus *Confusion Matrix* yang memiliki 2 kelas perhitungan. *Confusion Matrix multiclass* di gunakan untuk menemukan nilai *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan

False Negative (FN) untuk setiap kelas (negatif, netral, dan positif)[19]. Lebih jelasnya bisa dilihat pada tabel 2.1:

Tabel 2. 1 Confusion Matrix

Confusion Matrix		Prediksi		
		Negatif	Netral	Positif
Kelas	Negatif	TN	FNt	FP
	Netral	FN	TNt	FP
	Positif	FN	FNt	TP

Keterangan:

- True Positive* (TP), adalah nilai kelas yang diprediksi memiliki nilai positif dan faktanya positif.
- False Positive* (FP), adalah nilai kelas yang diprediksi memiliki nilai positif dan faktanya negatif atau netral.
- True Negative* (TN), adalah nilai kelas yang diprediksi memiliki nilai negatif dan faktanya negatif.
- False Negative* (FN), adalah nilai kelas yang diprediksi memiliki nilai negatif dan faktanya positif atau netral.
- True Neutral* (TNt) adalah nilai kelas yang diprediksi memiliki nilai netral dan faktanya netral.
- False neutral* (FNt) adalah nilai kelas yang diprediksi memiliki nilai netral dan faktanya positif dan negatif.

Empat hasil perhitungan menggunakan *Confusion Matrix* untuk pengujian meliputi :

- a. Menghitung akurasi, berikut rumus untuk menghitung akurasi:

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

- b. Menghitung presisi, berikut rumus untuk menghitung presisi:

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

- c. Menghitung *Recall*, berikut rumus untuk menghitung *Recall*:

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

- d. Menghitung *F1-score*, berikut rumus untuk menghitung *F1-score*:

$$F1 = 2x \frac{precision \times recall}{precision + recall} \quad (2.5)$$

2.2.6 Python

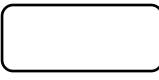
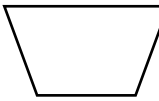
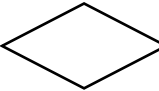
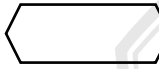
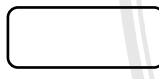

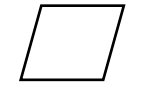
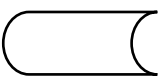

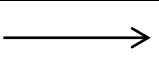


Python adalah salah satu bahasa pemrograman yang dapat digunakan untuk berbagai macam tugas, termasuk analisis data, *machine learning*, *data science*, *artificial intelligence* (AI), dan pengembangan situs web. *Python* menawarkan fitur-fitur lain selain sifatnya yang *general purpose*. Bahasa ini dikategorikan sebagai bahasa tingkat tinggi, berorientasi pada objek, dan ditulis secara dinamis[20].

2.3 Pemodelan

Pemodelan penelitian ini menggunakan *flowchart*, yang menggambarkan langkah-langkah dan urutan desain program untuk menghasilkan representasi grafis dari program itu sendiri. Pemrosesan dan tugas-tugas manual biasanya

membutuhkan flowchart, yang juga merupakan kumpulan simbol yang berfungsi untuk membangun. Berikut ini adalah contoh simbol-simbol pada flowchart[21]:

Tabel 2. 2 Simbol Flowchart

NO	Simbol	Nama	Keterangan
1		Proses	Sebuah fungsi yang menyebabkan perubahan data dalam proses
2		Simbol Manual	Pernyataan suatu tindakan manual yang tidak dilakukan menggunakan komputer
3		<i>Decision/logika</i>	Pemilihan opsi tertentu dengan “tidak” dan “ya”
4		<i>Prodefined Proses</i>	Memberikan harga awal pada penyediaan suatu tempat penyimpanan
5		Terminal	Memulai atau mengakhiri suatu program
6		<i>Offline Storage</i>	Menyimpan data pada suatu media tertentu
7		<i>Input/Output</i>	Untuk Proses <i>Input</i> atau <i>output</i>
8		<i>Disk Storage</i>	Menyatakan <i>input</i> berasal dari <i>disk</i> atau <i>output</i> disimpan dalam <i>disk</i>
9		<i>Documen</i>	Untuk mencetak dokumen
10		<i>Arus/flow</i>	Menentukan jalannya suatu proses
11		<i>Connector</i>	Bersambungnya antar proses dengan yang lainnya dalam satu halaman sama
12		<i>Offline Connector</i>	Menyambungkan suatu proses dengan proses lainnya dalam halaman berbeda

2.4 Perangkat Lunak yang Digunakan

2.4.1 *Microsoft Excel*

Microsoft Excel merupakan salah satu paket aplikasi *Microsoft Office* berupa aplikasi *spreadsheet* yang digunakan untuk mengolah data. Data tersebut dapat berupa teks (huruf), alfa numerik, atau numerik (angka), dan biasanya dikumpulkan dalam sebuah tabel untuk memberikan informasi yang berguna untuk mengambil keputusan terhadap suatu masalah atau kasus, dari yang sederhana hingga yang kompleks[22].

2.4.2 *TikTok*

Sebuah perangkat lunak yang disebut *TikTok* dengan menggunakan ponsel untuk menangkap dan berbagi momen unik dan tak ternilai dari seluruh dunia. Siapa pun dapat menjadi kreator berkat *TikTok*, yang juga mendorong orang untuk menyumbangkan kreasi artistik dalam video pendek berdurasi 15-60 detik. Kesederhanaan dan kenyamanan penggunaan program yang menghibur ini memungkinkan semua orang untuk menjadi kreator. Karena tidak membutuhkan banyak waktu, pengguna menemukan aplikasi yang berfokus pada video pendek lebih diminati dari sudut pandang pemasaran. Pengguna dapat dengan mudah menonton materi yang dipilih dan melanjutkan ke video lain jika tidak tertarik[23].

2.4.3 *Rapid Miner*

Rapid Miner, sebuah program sumber terbuka (*open source*), digunakan untuk menganalisis *text mining*, *data mining*, dan analisis prediksi. Untuk membantu pengguna membuat keputusan terbaik, aplikasi ini menggunakan berbagai pendekatan deskriptif dan prediktif. Terdapat sekitar 500 operator *data*

mining di *Rapid Miner*, termasuk operator untuk *input*, *output*, pemrosesan data, dan visualisasi. *Rapid Miner* adalah program yang berdiri sendiri untuk analisis data dan dapat diintegrasikan ke dalam produknya sendiri sebagai mesin *data mining*. Program ini ditulis dalam bahasa Java sehingga dapat berjalan di semua sistem operasi. *Graphic User Interface* (GUI) yang ditawarkan oleh *Rapid Miner* memfasilitasi pembuatan pipeline analitik. Sebuah *file XML* (*Extensible Markup Language*) yang mendefinisikan prosedur analisis yang ingin diterapkan oleh pengguna pada data akan dibuat oleh GUI ini. *Rapid Miner* kemudian membaca *file* ini dan melakukan analisis secara otomatis[24].

2.4.4 *Data Miner*

Data Miner merupakan Ekstensi peramban *Google Chrome* dan *Edge* yang memfasilitasi ekstraksi dan pengambilan data dari situs *web* dan menyimpannya dalam *file CSV* atau *spreadsheet Excel*. *Data Miner* memiliki UI intuitif untuk membantu menjalankan ekstraksi data tingkat lanjut dan pengkikisan web. *Data Miner* *mengscraping* satu halaman atau menjelajahi situs dan mengekstrak data dari beberapa halaman seperti hasil pencarian, produk dan harga, informasi kontak, email, nomor telepon, dan banyak lagi. Setelah itu, data di *scraping* dan dikonversi oleh *Data Miner* ke dalam format *file CSV* atau *Microsoft Excel* yang bersih. *Data Miner* mengumpulkan informasi dari halaman web yang muncul di peramban secara otomatis ketika Anda mengklik tombol dan tautan serta mengikuti sub halaman, dan membuka *pop-up*. dan juga mengekstrak teks dari halaman-halaman web tersebut[25].

2.4.5 *Visual Studio Code*

Visual Studio Code merupakan editor teks yang ringan dan kuat yang dikembangkan oleh *Microsoft* untuk sistem operasi multiplatform, sehingga tersedia juga dalam versi *Windows, Linux, dan Mac*. Selain *Javascript, Typescript*, dan *Node.js*, editor teks ini mendukung bahasa pemrograman lain termasuk *C++*, *C#*, *Python*, *Go*, *Java*, *PHP*, dan lainnya melalui *plugin* yang dapat ditambahkan melalui pasar *Visual Studio Code*[26].



BAB III

ANALISIS PERANCANGAN

3.1 Gambaran Umum Objek Penelitian

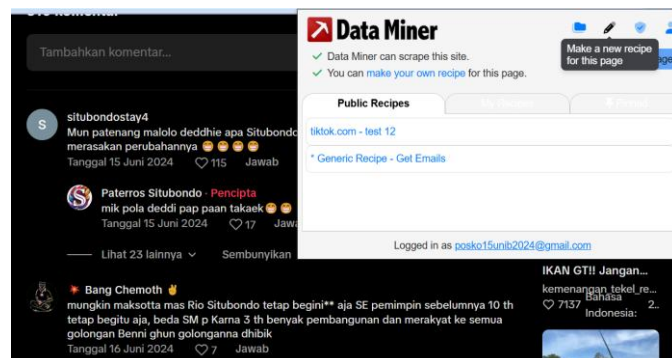
Media sosial *platform TikTok* adalah tempat di mana pengguna dapat membuat, mengedit, dan berbagi konten singkat yang berpusat pada video pendek dan biasanya memiliki musik atau efek artistik. *TikTok* terkenal dengan algoritmanya yang canggih, dapat mengubah konten sesuai dengan preferensi pengguna sehingga meningkatkan keterlibatan dan interaksi. *Platform* ini telah berkembang menjadi tempat bagi sejumlah kelompok, seperti perusahaan, institusi pendidikan, dan kreator, yang menggunakannya untuk pemasaran, pembelajaran, dan hiburan. *TikTok* sebuah fenomena digital dengan miliaran unduhan di seluruh dunia, berdampak pada taktik pemasaran, komunikasi, dan tren budaya di era modern. Objek pada penelitian ini adalah komentar terkait pemilu di *platform TikTok*, yang menyediakan wadah bagi orang-orang untuk berkomunikasi dan berbagi pandangan politik mereka. Sebagai *platform* media sosial yang berfokus pada video pendek, *TikTok* memungkinkan pengguna untuk terlibat dalam wacana politik melalui fungsi komentarnya, yang sering kali mewakili berbagai sudut pandang, mulai dari dukungan dan kritik hingga penyebaran informasi dan informasi yang salah. Sikap politik yang muncul, tren *viral*, dan algoritma *platform*, semuanya dapat berdampak pada pola pemikiran *netizen*.

3.1.1 *Scrapping Data TikTok*

Data yang digunakan dalam penelitian ini menggunakan data sekunder berupa komentar di *platform TikTok*, yang berkaitan dengan PILKADA di

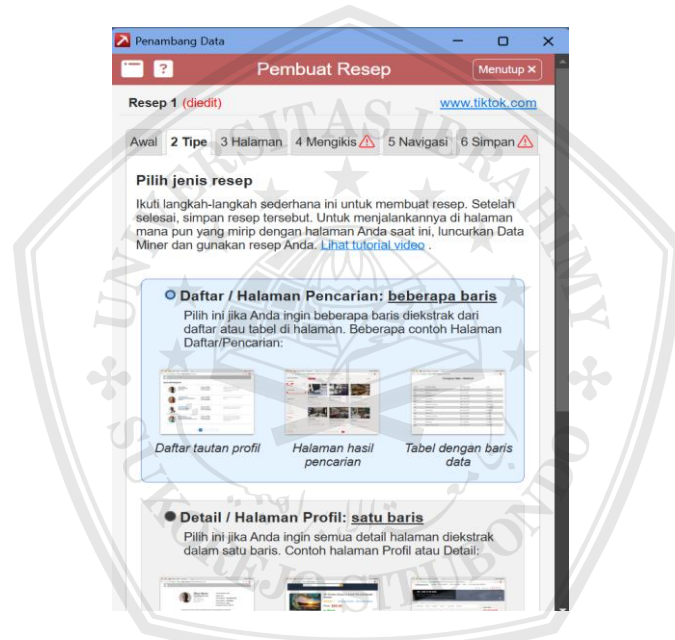
Kabupaten Situbondo. Data tersebut diperoleh dengan cara melakukan *scrapping* pada komentar terkait PILKADA di Kabupaten Situbondo pada *platform TikTok*, memanfaatkan perangkat lunak *Data Miner*. *Scrapping data* dimulai dengan memasukkan kata kunci yang berhubungan dengan penelitian tersebut di aplikasi *TikTok*, kemudian mencari video *TikTok* yang memiliki banyak respon dari *netizen*. Data yang telah dikumpulkan melalui *scrapping* kemudian diolah menggunakan *text mining* untuk membersihkan dan mempersiapkan pada analisis lebih lanjut. Data yang berhasil didapatkan dalam penelitian ini sebanyak 494 data yang hasil *outputnya* berupa data *excel*. Berikut adalah langkah-langkah dari proses *scrapping data TikTok* :

- Mengumpulkan beberapa video yang berkaitan dengan PILKADA.
- Menyiapkan halaman video yang telah dikumpulkan dan akan dilakukan *scrapping*.
- Membuka *Data Miner* pada ekstensi *Chrome*.
- Membuat resep baru untuk meng-*Scrapping page TikTok* seperti pada gambar 3.1.



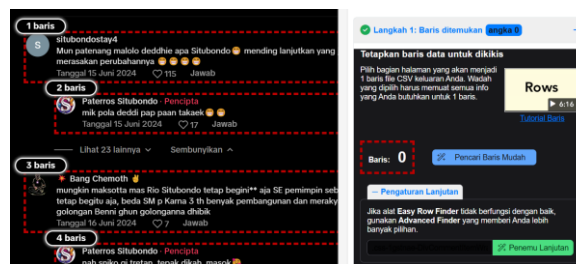
Gambar 3. 1 Buat Resep Data Miner

- e. Selanjutnya memilih tipe jenis resep yang akan di buat. Ada 3 jenis resep yaitu daftar/halaman pencarian jika ingin meng-*scrapping* beberapa baris pada halaman *web*, jenis resep detail/halaman profil jika ingin meng-*scrapping* detail halaman *web* dalam satu baris, dan jenis resep mengisi formulir dan mengikis jika ingin mengisi formulir dan kemudian mengiklankan hasilnya. Pada gambar 3.2 jenis resep daftar/halaman di pilih untuk meng-*scrapping* komentar *TikTok*



Gambar 3. 2 Pilih Jenis Resep

- f. Menetapkan baris yang akan di lakukan *scrapping* pada halaman yang telah disediakan. Seperti yang di tunjukkan pada gambar 3.3.



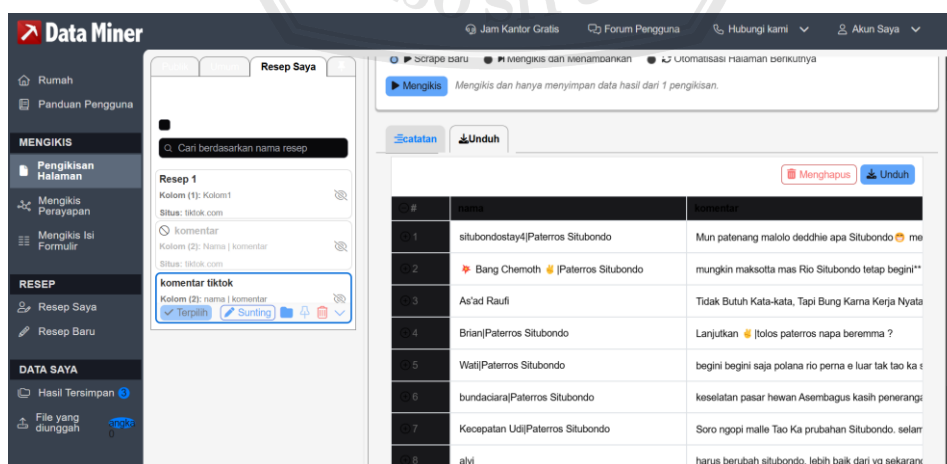
Gambar 3. 3 Menetapkan Baris

- g. Menetapkan kolom dan menentukan jenis nama kolom yang akan digunakan. Seperti pada gambar 3.4.



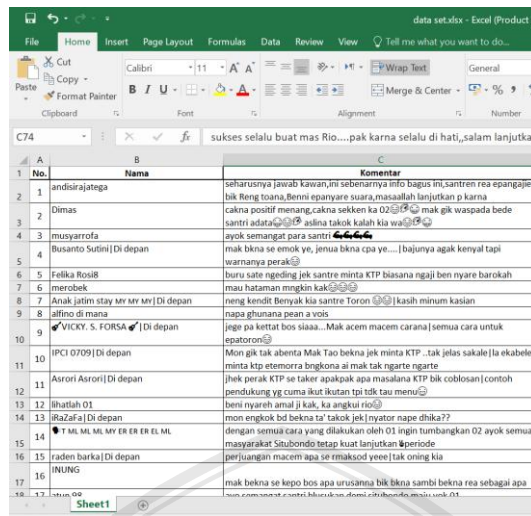
Gambar 3. 4 Menetapkan Kolom dan Menentukan Nama Jenis

- h. Menyimpan hasil *scrapping* TikTok seperti pada gambar 3.5 dapat di simpan dalam bentuk *file excel* atau *csv*.



Gambar 3. 5 Hasil Scrapping

- i. Hasil *scrapping* yang disimpan pada *Microsoft Office Excel*



No	Nama	Komentar
1	andisirajatega	seharusnya jawab kawan, ini sebenarnya info bagus ini, santrén rea epangaje bik Reng toana, Benni epanyare suara, masalah lanjutan p karna
2	Dimas	cahna postif menang, cahna sekean ka 02 mak gik waspada bede santri sadat, adina takok kalah kua wa
3	musyarrofa	ayok semangat para santri
4	Busanto Sutini Di depan	mak bkna se emok ye, jenua bkna cpa ye... bajunya agak kenyal tapi warnanya perak
5	Falika Rosih	buru sate ngeding jek santrén minta KTP biasana ngaji ben nyare barokah
6	merobek	mau butaman ngelén tak
7	Anak jatim stay Mr My My Di depan	neng kendit Benyak kua santrén Toron kash minum kasian
8	alfino di mana	napa ghunana pean a vois
9	VICKY. S. FORSA Di depan	jega pa kettat bos siaaa... Mak acem macem carana semua cara untuk eparotoni
10	IPCI 0709 Di depan	Moni gik tak abenta Mak Tao bekna jek minta KTP tak jelas sakale la ekabele minta ktp etemorra bngkona ai mak tak ngarte ngarte
11	Asrori Asrori Di depan	Jhek perak KTP se taker apakpak apa masalaha KTP bik coblosan contoh pendukung yg cuma ikut ikutan tpi tdk tau menu
12	ihattah 01	beni nyarah amal ji kak, ka angku rihi
13	ihazZaf Di depan	mon enguk bd bekna ta' talok jek nyator nape dhika??
14	T M L M L M Y ER ER EL M L	dengan semia cara yang dilakukan oleh 01 ingin tumbangkan 02 ayok semua masyarakat Situbondo tetap kuat lanjutan periode
15	raden barka Di depan	perjuangan macem apa se rmaksod yeee tak oning kua
16	INUNG	mak bekna se kepo bos apa urusama bik bkna sambi bekna rea sebagai apa
17		

Gambar 3. 6 Hasil Berbentuk File excel

Dari gambar 3.6 data yang diperoleh terdapat beberapa atribut yakni nama dan komentar, data yang telah diperoleh akan dilakukan *preprocessing data* untuk menyeleksi data teks agar lebih terstruktur.

3.2 Alur Proses

Alur proses yang merupakan gambaran prosedur yang akan digunakan untuk memecahkan masalah, secara rinci proses ini akan menjelaskan mengenai tahapan sebuah penelitian. Prosedur ini akan memudahkan penelitian untuk berjalan sesuai dengan yang diinginkan.

3.2.1 Identifikasi dan Analisis proses Bisnis

Pada bagian ini dilakukan identifikasi terkait dengan objek penelitian yang sedang dilakukan yaitu data komentar *TikTok* pada PILKADA di Kabupaten Situbondo, pada tahap ini akan dilakukan beberapa proses yang terkait sehingga data siap digunakan :

- a. *Preprocessing data*
- b. *Labeling data*
- c. *Metode Naive bayes*

3.2.2 Analisis Proses Bisnis

Analisis Proses Bisnis ini terdiri dari beberapa penjelasan tentang proses bisnis yaitu *Preprocessing data* dan metode *Naive Bayes* :

- a. *Preprocessing data*

Sebelum proses klasifikasi, langkah yang dikenal sebagai *Preprocessing*, bertujuan untuk mengubah data yang belum diproses menjadi data berkualitas tinggi yang dapat digunakan untuk langkah pemrosesan lebih lanjut. Untuk menghilangkan data yang tidak konsisten atau bermasalah, langkah ini diterapkan pada data mentah. data yang bermasalah, seperti data yang berisi nol atau *error*[27]. Tahapan dari *Preprocessing Data* meliputi *Cleansing*, *Case Folding*, *Tokenizing*, *Remove Stopword*, dan *Stemming Data* :

1. *Cleansing*

Cleansing adalah proses menghilangkan data atau teks yang kosong atau tidak perlu. Prosedur pembersihan sangat membantu untuk memverifikasi bahwa data analitik sudah akurat dan bersih[28].

2. *Case Folding*

Teknik mengubah huruf kapital dalam kumpulan data menjadi huruf kecil dikenal sebagai *Case Folding*. Dalam pemrosesan

teks, teknik ini sering digunakan untuk menyamakan format huruf untuk analisis yang lebih sederhana[28].

3. *Tokenizing*

Frasa akan dibagi menjadi unit kata, atau token, selama proses *Tokenizing*, dan ini akan dipisahkan dengan koma atau tanda baca lainnya[28].

4. *Remove Stopword*

Remove Stopword merupakan proses di mana kata-kata penghubung yang dianggap tidak perlu dan tidak mempengaruhi emosi kalimat akan dihilangkan dari analisis[28].

5. *Stemming Data*

Stemming Data merupakan proses memetakan varian kata ke kata sumbernya. Pengguna umumnya sebagai dasar dari metode pemrosesan bahasa alami untuk pencarian informasi yang sukses dan efisien. Proses *stemming* melibatkan identifikasi kata dasar suatu istilah[29].

b. Metode *Naive Bayes*

Berikut merupakan tahapan pengujian algoritma *Naive Bayes*

Clasifier di bawah ini:

1. Menghitung probabilitas dari V_j

$$P(v_j) = \frac{|doc\ j|}{|training|} \quad (3.1)$$

Pada data *training* dengan persamaan (3.1) di mana V_j merupakan Kategori sentimen, yaitu $V_1 =$ negatif, $V_2 =$ positif dan $V_3 =$ netral[30].

2. Menghitung probabilitas kata a_i

$$P(a_i|v_j) = \frac{ni + 1}{|n + \text{kosakata}|} \quad (3.2)$$

Pada kategori V_j dengan persamaan (3.2)

3. Model probabilitas NBC disimpan dan digunakan untuk tahap testing.
4. Menghitung probabilitas tertinggi dari kategori sentimen yang diujikan (V_{map}) dengan persamaan (2.1)
5. Mencari nilai V_{map} paling maksimum dan masukkan komentar tersebut pada kategori dengan V_{map} maksimum (3.3).

$$V_{MAP} = \text{argmaz } P(v_j) \prod_i P(a_i|v_j) \quad (3.3)$$

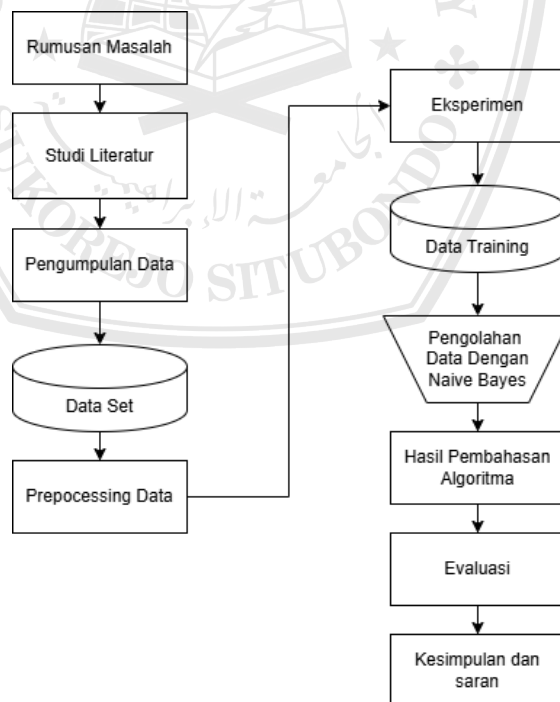
c. *Confusion Matrix*

Menghitung ketepatan klasifikasi dari algoritma *Naive Bayes Classifier* berdasarkan hasil *Confusion Matrix* yang di peroleh, berikut tahapan perhitungannya:

1. Menghitung nilai akurasi, atau proporsi kelas yang diprediksi secara akurat oleh model yang sudah dibuat (2.2).
2. Menentukan nilai presisi, atau tingkat keakuratan antara informasi yang diminta pengguna dan respons sistem (2.3).

3. Menentukan nilai *Recall*, sering disebut sebagai sensitivitas, yang merupakan proporsi data yang diprediksi oleh program daripada kelas yang sebenarnya (2.4).
 4. Menentukan nilai F1-Score, salah satu formula penilaian dalam data yang mengintegrasikan akurasi dan recall. Rata-rata tertimbang recall dan akurasi dibandingkan dalam nilai F1-Score (2.5).
- d. Alur Tahapan Algoritma *Naive Bayes*

Langkah-langkah yang dilakukan dalam analisis sentimen pada data komentar *TikTok* bisa dilihat pada gambar 3.6 sebagai berikut:



Gambar 3. 7 Tahapan Algoritma *Naive Bayes*

3.2.3 Identifikasi dan Analisis Kebutuhan

a. Identifikasi Kebutuhan Fungsional

Kebutuhan fungsional memiliki peran penting dalam sebuah penelitian karena mereka memastikan bahwa proses berjalan dengan lancar.

b. Analisis Kebutuhan Fungsional

Pada tahap ini data akan diolah menggunakan sebuah tahapan data yakni *Preprocessing* data. Data diolah melalui enam tahapan yaitu *Cleaning*, *Case Folding*, *Tokenizing*, *Remove Stopword*, dan *Stemming Data*. *Preprocessing* dilakukan agar data yang didapat setelah proses *Scrapping* data komentar *TikTok* yang merupakan data mentah, setelah dilakukan *preprocessing* akan menjadi data yang siap digunakan untuk analisis sentimen menggunakan metode yang telah ditentukan, sehingga data tidak mengalami *error* karena adanya data yang tidak memiliki nilai. Setelah proses *preprocessing* selesai dilakukan data akan melewati proses pelabelan dalam penelitian ini ada 3 label yang digunakan yakni label yang bernilai positif, negatif dan netral. Setelah data melewati proses pelabelan data akan dibagi menjadi *data training* dan *data testing*, hal ini dilakukan untuk memudahkan proses implementasi metode algoritma yang akan digunakan yakni algoritma *Naive Bayes*. Pada proses perhitungan menggunakan algoritma *Naive Bayes* akan mencari nilai probabilitas

tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat.

3.2.4 Identifikasi dan Analisis Alternatif Solusi

a. Identifikasi Alternatif Solusi

Proses ini berfungsi untuk menganalisis alternatif yang digunakan dalam pembuatan program. Terdapat dua alternatif solusi dalam penelitian ini. Tabel 3.1 merupakan tabel identifikasi alternatif solusi.

Tabel 3. 1 Identifikasi Alternatif Solusi

Karkteristik	Alternatif 1 (Python)	Alternatif 2 (RapidMiner)
Bagian sistem yang terkomputerisasi.	Semua kebutuhan fungsional akan terpenuhi	Semua kebutuhan fungsional akan terpenuhi.
Keuntungan, deskripsi ringkas keuntungan yang akan direalisasikan	Pengembangan lebih cepat dan mudah dilakukan bagi pemula	Proses lebih cepat bagi pemula
Perangkat lunak aplikasi	perangkat lunak yang akan digunakan adalah <i>Visual Studio Code</i> menggunakan bahasa <i>Python</i>	Perangkat lunak yang digunakan yang akan digunakan adalah GUI, Java API dan <i>data mining</i> .
Metode pemrosesan data, pada biasanya kombinasi <i>online</i> , batch, deferred, remote batch dan reel time	Editor teks yang dikembangkan oleh Microsoft untuk sistem operasi multiplatform	
Alat <i>Output</i>	Monitor dan Printer	Monitor dan Printer
Alat <i>Input</i>	<i>Keyboard</i> dan <i>Mouse</i>	<i>Keyboard</i> dan <i>Mouse</i>
Alat penyimpanan	Data disimpan dalam folder ruang kerja (.vsWork)	Data disimpan dalam tabel repositori

b. Analisis Kelayakan Alternatif Solusi

Proses ini berfungsi untuk menganalisis sistem yang akan dibuat berdasarkan dokumen yang didapatkan sebelumnya. Tabel 3.2 menjelaskan identifikasi kelayakan alternatif.

Tabel 3. 2 Analisis Kelayakan Alternatif Solusi

Karakteristik	Bobot	Alternatif 1 (Python)	Alternatif 2 (RapidMiner)
Kelayakan Operasional Fungsional	40%	Mendukung seluruh kebutuhan fungsional serta pengembangan yang lebih mudah dan bisa dikerjakan oleh banyak orang	Mendukung seluruh kebutuhan fungsional, mudah dilakukan oleh pemula.
Skor		40	30
Kelayakan teknis teknologi	30%	Teknologi yang digunakan sudah sangat mencukupi dalam keahlian	Teknologi yang digunakan dalam manajemen sudah memadai dan sudah cukup.
Skor		25	30
Kelayakan ekonomis biaya pengembangan	30%	Pengembangan sistem tidak memerlukan biaya dalam pengoperasiannya	Tidak ada pengembangan dan bersifat komersial
Skor		30	20
Total	100%	95	80

3.2.5 Desain Proses

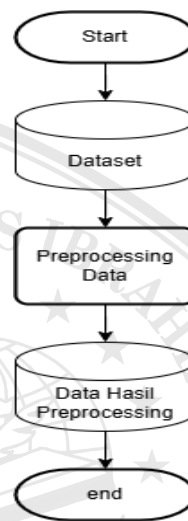
Desain proses merupakan tahapan dalam merancang sebuah sistem. Tahapan ini dilakukan agar proses sistem berjalan sesuai dengan alur yang diinginkan.

a. Identifikasi Desain Proses

Pada identifikasi desain proses menjelaskan tentang alur dari analisis yang dilakukan. Bagian ini fokus pada proses yang akan dilakukan yakni proses *input* data, perhitungan dengan metode dan evaluasi hasil dari metode. Berikut adalah desain proses yang akan dilakukan.

1. *Preprocessing Data*

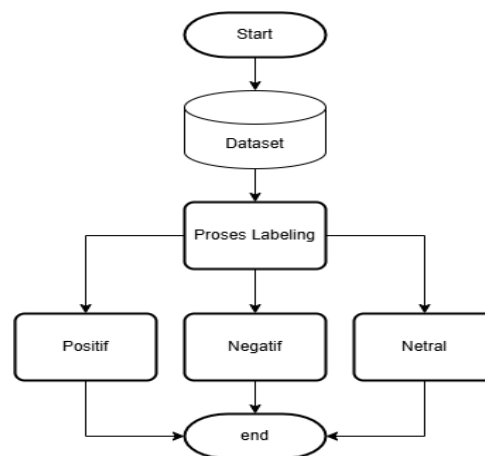
Proses ini dilakukan untuk mengolah data menjadi sebuah dataset. Data diperoleh dari hasil *Scrapping* komentar *TikTok* menghasilkan *output* dataset berupa *file excel* lalu di *input*-kan dan siap dilakukan proses perhitungan dengan algoritma. Gambar 3.7 menjelaskan tentang alur proses *Preprocessing data*.



Gambar 3. 8 Preprocessing Data

2. *Labeling Data*

Input data yang sudah dilakukan *preprocessing*, dan selanjutnya dilakukan proses *labeling*, sentimen positif, negatif dan netral. Gambar 3.8 merupakan alur proses *labeling data*.

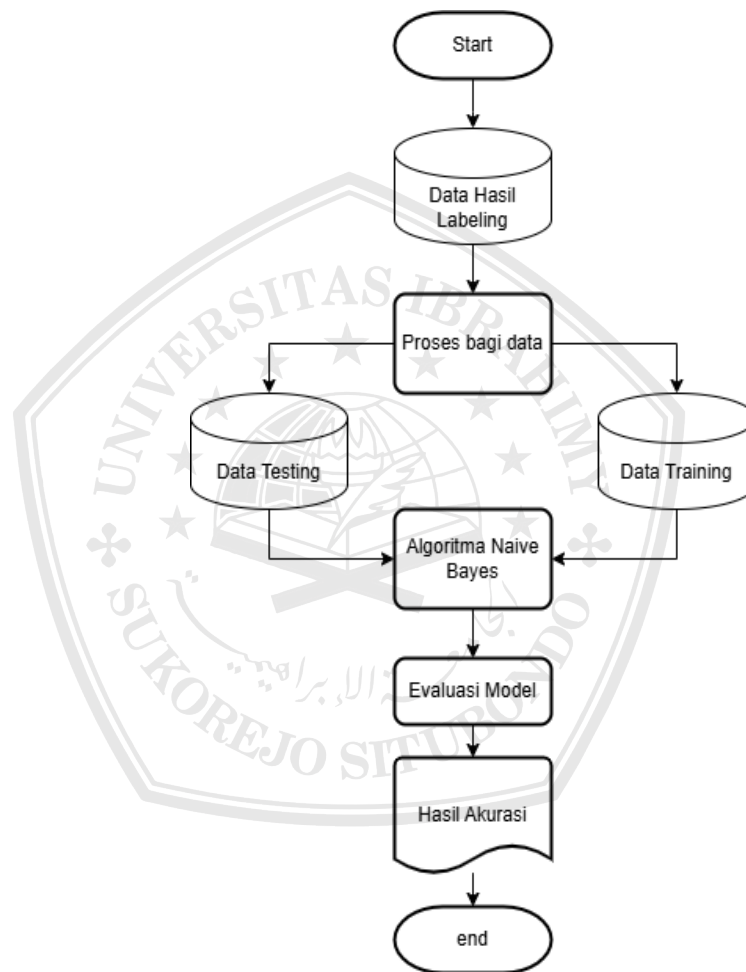


Gambar 3. 9 Alur Labeling Data

3. Metode *Naive Bayes*.

Input data hasil *labeling*, bagi menjadi *data testing* dan *data training* untuk menghitung akurasi, di *input* data *excel* dari hasil *labeling data*. dilakukan eksperimen melakukan algoritma *Naive Bayes*. *Output* yang dihasilkan adalah Nilai hasil akurasi.

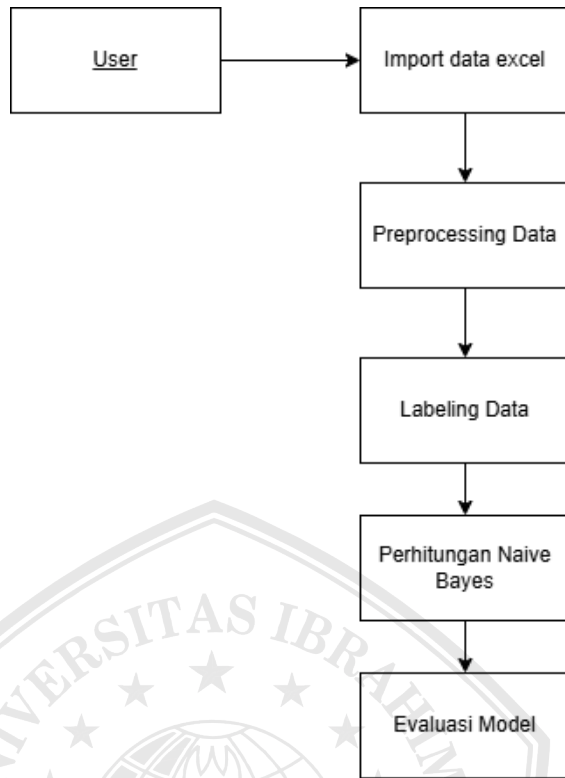
Gambar 3.9 menjelaskan alur metode *Naive Bayes*



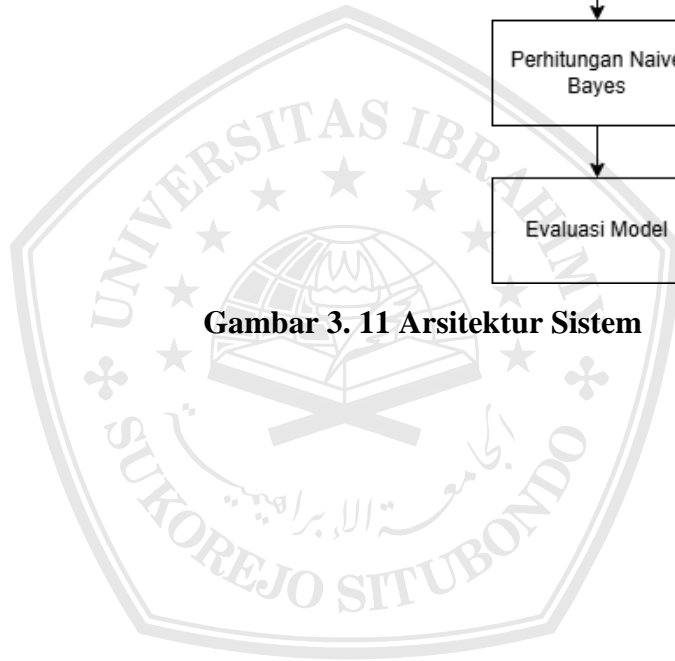
Gambar 3. 10 Alur Metode *Naive Bayes*

b. Arsitektur Sistem

Arsitektur sistem berfungsi sebagai penjelasan dari pola sistem analisis yang akan dilakukan, berikut adalah arsitektur sistem yang akan dilakukan. Gambar 3.10 menjelaskan arsitektur sistem.



Gambar 3. 11 Arsitektur Sistem



BAB IV

HASIL DAN PEMBAHASAN

4.1 Kebutuhan Sistem

Dalam bahasa pemrograman Python sistem dapat dijalankan dengan memperhatikan beberapa perangkat yang bisa menunjang keberhasilan sistem. Berikut adalah perangkat yang digunakan.

a. *Software*

Berikut ini adalah *Software* yang digunakan sebagai pendukung pada sistem:

1. *Visual Studio Code*
2. *Data Miner*
3. *Microsoft excel*

b. *Hardware*

Berikut ini adalah *Hardware* yang digunakan sebagai pendukung pada sistem:

1. Laptop Asus
2. *Mouse dan Keyboard*
3. *Processor Intel Core i3*
4. *Hard Disk dengan kapasitas 1 Terabyte (TB)*
5. *RAM 8 GB*

4.2 Preprocessing Data

Sebelum masuk pada tahap *preprocessing* data, dilakukan proses menterjemah komentar *netizen* yang berbahasa Madura ke dalam bahasa Indonesia. Beberapa tahapan yang dilakukan pada *preprocessing* antara lain:

a. *Cleaning*

Cleaning merupakan tempat di mana prosedur *preprocessing* dimulai. Untuk meminimalkan gangguan atau kesalahan selama pemrosesan data, karakter non alfabet dibersihkan atau dihilangkan pada langkah ini. *whitespace*, *emoticon*, *link double*, *URL*, *mention*, duplikasi teks dan angka duplikat semuanya dihapus pada tahap ini. Fungsi pustaka NLTK (*Natural Language Toolkit*) dari bahasa *Python* yang digunakan dalam prosedur ini. Pada tabel 4.1 menunjukkan teks sebelum dan sesudah proses *cleaning*.

Tabel 4. 1 Proses *Cleaning*

Komentar	Cleaning
seharusnya jawab kawan,ini sebenarnya info bagus ini,santren rea epangajie bik Reng toana,Benni epanyare suara,masaallah lanjutkan p karna	seharusnya jawab kawan ini sebenarnya info bagus ini Santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna
cakna positif menang,cakna sekken ka 02🤔🤔🤔 mak gik waspada bede santri adata🤔🤔🤔 aslina takok kalah kia wa🤔🤔🤔	katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslinya takut kalah juga ya
ayok semangat para santri 🤔🤔🤔	ayo semangat para santri
mak bkna se emok ye, jenua bkna cpa ye....	kok kamu yang repot ya emangnya kamu siapa
buru sate ngeding jek santre minta KTP biasana ngaji ben nyare barokah	baru sekarang dengar santri minta KTP biasanya ngaji dan cari barokah

Pada segmen program 4.1 merupakan proses *Cleaning* untuk menghapus data atau kata yang tak memiliki nilai.

Segmen Program 4. 1 *Cleaning*

```

1 import re
2 import string
3 import nltk
4 # Fungsi pembersih
5 def remove_emojis(tweet):
6     if tweet is not None and isinstance(tweet, str):

```

Segmen Program 4.1 (Lanjutan)

```
7 emoji_pattern = re.compile(
8     "["
9     u"\U0001F600-\U0001F64F"
10    u"\U0001F300-\U0001F5FF"
11    u"\U0001F680-\U0001F6FF"
12    u"\U0001F1E0-\U0001F1FF"
13    u"\U00002702-\U000027B0"
14    u"\U000024C2-\U0001F251"
15    "]+", flags=re.UNICODE
16 )
17
18     return emoji_pattern.sub(r'', tweet)
19 else:
20     return tweet
21
22 def remove_symbols(tweet):
23     if tweet is not None and isinstance(tweet, str):
24         return re.sub(r'^[a-zA-Z0-9\s]', '', tweet)
25     return tweet
26
27 def remove_numbers(tweet):
28     if text is not None and isinstance(tweet, str):
29         tweet=re.sub(r'\d+', '', tweet)
30     return tweet
31
32 def remove_username(text):
33     if text is not None and isinstance(text, str):
34         return re.sub(r'@[^\s]+', '', text)
35     else:
36         return text
37
38 # Pipeline pembersih
39 df['cleaning'] = df['terjemahan'].apply(remove_emojis)
40 df['cleaning'] = df['cleaning'].apply(remove_symbols)
41 df['cleaning'] = df['cleaning'].apply(remove_numbers)
42 df['cleaning'] = df['cleaning'].apply(remove_username)
43
44 (df.head(5))
```

b. Case Folding

Terlepas dari apakah huruf sebelumnya ditulis dengan huruf besar atau tidak, *case folding* adalah prosedur *preprocessing data* yang berusaha mengubah

semua karakter menjadi huruf kecil. Pada tahap ini dengan memanfaatkan fungsi `.lower()` dari pustaka `pandas` teks akan tampil dalam format huruf kecil yang seragam setelah proses ini dijalankan. Tabel 4.2 merupakan proses sebelum dan sesudah *case folding*, dan pada segmen program 4.2 merupakan proses *case folding*.

Tabel 4. 2 Case Folding

Cleaning	Case Folding
seharusnya jawab kawan ini sebenarnya info bagus ini Santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna	seharusnya jawab kawan ini sebenarnya info bagus ini santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna
katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslanya takut kalah juga ya	katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslanya takut kalah juga ya
ayo semangat para santri	ayo semangat para santri
kok kamu yang repot ya emangnya kamu siapa	kok kamu yang repot ya emangnya kamu siapa
baru sekarang dengar santri minta KTP biasanya ngaji dan cari barokah	baru sekarang dengar santri minta ktp biasanya ngaji dan cari barokah

Segmen Program 4. 2 Case Folding

```

1 #CASE FOLDING
2 def case_folding(text):
3     if isinstance(text, str):
4         return text.lower()
5     else:
6         return text
7
8 df['case_folding'] = df['cleaning'].apply(case_folding)
9 (df.head(5))

```

c. *Tokenizing*

Kalimat-kalimat kini diurai menjadi perkata yang disebut proses *Tokenizing*.

Untuk memudahkan analisis pada langkah pemrosesan berikutnya, tujuan

utamanya adalah memisahkan setiap kata yang membentuk sebuah frasa. Dalam proses ini memanfaatkan fungsi dari `.split()` atau NLTK dan perpustakaan NLP lainnya. Pada tabel 4.3 merupakan proses sebelum dan sesudah *Tokenizing*, dan pada seegmen program 4.3 merupakan proses dari *Tokenizing*.

Tabel 4.3 Tokenizing

Case Folding	Tokenizing
seharusnya jawab kawan ini sebenarnya info bagus ini santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna	['seharusnya', 'jawab', 'kawan', 'ini', 'sebenarnya', 'info', 'bagus', 'ini', 'santri', 'ini', 'disuruh', 'ngaji', 'oleh', 'orang', 'tuanya', 'bukan', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']
katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslanya takut kalah juga ya	['katanya', 'positif', 'menangkatanya', 'sudah', 'mantap', 'ke', 'kosong', 'dua', 'kok', 'masih', 'khawatir', 'ada', 'santri', 'men', 'dataaslanya', 'takut', 'kalah', 'juga', 'ya']
ayo semangat para santri	['ayo', 'semangat', 'para', 'santri']
kok kamu yang repot ya emangnya kamu siapa	['kok', 'kamu', 'yang', 'repot', 'ya', 'emangnya', 'kamu', 'siapa']
baru sekarang dengar santri minta ktp biasanya ngaji dan cari barokah	['baru', 'sekarang', 'dengar', 'santri', 'minta', 'ktp', 'biasanya', 'ngaji', 'dan', 'cari', 'barokah']

Segmen Program 4.3 Tokenizing

```

1 #TOKENISASI
2 import nltk
3 nltk.download('punkt')
4 from nltk.tokenize import word_tokenize
5 def tokenize(text):
6     if pd.isna(text):
7         return []
8     return str(text).split()
9
10 df['token'] = df['case_folding'].apply(tokenize)
11 (df.head(5))

```

d. *Remove Stopword*

Pada tahap ini melibatkan penghilangan kata-kata yang dianggap secara signifikan berkontribusi pada makna dokumen. Kata-kata yang bersifat umum dan tidak menyampaikan informasi yang tepat, seperti “ini,” “kok,” “masih,” “ke,” “atau,” dan sebagainya, biasanya dihilangkan. Untuk menyaring hanya istilah-istilah yang relevan, prosedur ini sering dilakukan setelah proses tokenisasi. Pada tabel 4.4 merupakan proses sebelum dan sesudah *Remove Stopword*.

Tabel 4. 4 *Remove Stopword*

Tokenizing	Remove Stopword
['seharusnya', 'jawab', 'kawan', 'ini', 'sebenarnya', 'info', 'bagus', 'ini', 'santri', 'ini', 'disuruh', 'ngaji', 'oleh', 'orang', 'tuanya', 'bukan', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']	['kawan', 'info', 'bagus', 'santri', 'disuruh', 'ngaji', 'orang', 'tuanya', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']
['katanya', 'positif', 'menangkatanya', 'sudah', 'mantap', 'ke', 'kosong', 'dua', 'kok', 'masih', 'khawatir', 'ada', 'santri', 'men', 'dataaslina', 'takut', 'kalah', 'ya', 'juga', 'ya']	['positif', 'menangkatanya', 'mantap', 'kosong', 'khawatir', 'santri', 'men', 'dataaslina', 'takut', 'kalah', 'ya']
['ayo', 'semangat', 'para', 'santri']	['ayo', 'semangat', 'santri']
['kok', 'kamu', 'yang', 'repot', 'ya', 'emangnya', 'kamu', 'siapa']	['repot', 'ya', 'emangnya']
['baru', 'sekarang', 'dengar', 'santri', 'minta', 'ktp', 'biasanya', 'ngaji', 'dan', 'cari', 'barokah']	['dengar', 'santri', 'ktp', 'ngaji', 'cari', 'barokah']

Pada segmen program 4.4 merupakan proses dari *Remove Stopword* untuk menyaring kata-kata yang tidak memiliki kontribusi signifikan terhadap makna dokumen.

Segmen Program 4. 4 Stopword Removal

```

1 #STOPWORD REMOVAL
2 from nltk.corpus import stopwords
3 nltk.download('stopwords')
4 stop_words = set(stopwords.words('indonesian'))
5
6 def remove_stopwords(tokens):
7     return [word for word in tokens if word.lower() not
8     in stop_words]
9
10 df['stopwprd removal'] = df['token'].apply(lambda x:
11 remove_stopwords(x))
12 (df.head(5))

```

e. *Stemming Data*

Tahap ini, yang dikenal sebagai stemming, merupakan tahap terakhir dalam prosedur persiapan teks. Tujuan utamanya adalah mengembalikan kata-kata yang terinfleksi ke bentuk dasarnya. Misalnya, kata-kata “bermain”, “memasak”, dan “terlambat” akan diubah menjadi “main”, ‘masak’ , dan “lambat”. Prosedur ini dilakukan setelah tahap tokenisasi dan penghapusan kata stopwords. Sastrawi adalah salah satu pustaka paling umum dalam bahasa Indonesia. Pada tabel 4.5 merupakan proses sebelum dan sesudah *Stemming Data*.

Tabel 4. 5 Stemming Data

Remove Stopword	Stemming Data
['kawan', 'info', 'bagus', 'santri', 'disuruh', 'ngaji', 'orang', 'tuanya', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']	kawan info bagus santri suruh ngaji orang tua suruh cari suara masaallah lanjut p karna
['positif', 'menangkatanya', 'mantap', 'kosong', 'khawatir', 'santri', 'men', 'dataaslinya', 'takut', 'kalah', 'ya']	positif menangkatanya mantap kosong khawatir santri men dataaslinya takut kalah ya
['ayo', 'semangat', 'santri']	ayo semangat santri

Tabel 4.5 Stemming Data Lanjutan

['repot', 'ya', 'emangnya']	repot ya emang
['dengar', 'santri', 'ktp', 'ngaji', 'cari', 'barokah']	dengar santri ktp ngaji cari barokah

Pada segmen program 4.5 merupakan proses dari *Stemming Data* menggunakan pustaka Sastrawi

Segmen Program 4.5 Stemming Data

```
1 #STEMMING
2 from Sastrawi.Stemmer.StemmerFactory import
3 StemmerFactory
4 from nltk.stem import PorterStemmer
5 from nltk.stem.snowball import SnowballStemmer
6
7 factory = StemmerFactory()
8 stemmer = factory.create_stemmer()
9
10 def stemming(tokens):
11     return [stemmer.stem(token) for token in tokens]
12
13 df['stemming data'] = df['stopwprd removal'].apply(lambda
14 x: ' '.join(stemming(x)))
15 (df.head(5))
```

4.3 Wordcloud

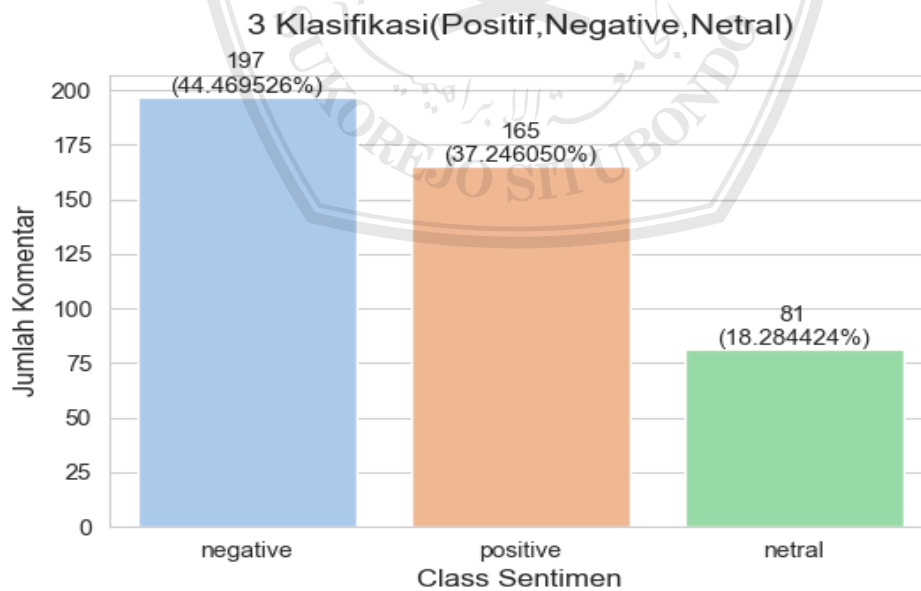
Sepertinya kata-kata umum seperti “yang”, “potong”, “tidak”, dan “lanjutkan” mendominasi frekuensi kata pada fase visualisasi *wordcloud* pertama, yang dilakukan sebelum tahap *preprocessing*. Karena teknik pembersihan teks seperti *cleaning*, *tokenizing*, *stemming* dan *removal stopword* belum diterapkan. Gambar 4.1 adalah *wordcloud* sebelum dilakukan *preprocessing data*. Sedangkan pada gambar 4.2 adalah frekuensi kata sebelum dilakukan *preprocessing data*.

membagi sentimen. Pada gambar 4.5 merupakan contoh dari data yang sudah diberi label.

	stemming data	Sentimen
0	kawan info bagus santri suruh ngaji orang tua suruh cari suara masaallah lanjut p karna	negative
1	positif menangkatanya mantap kosong khawatir santri men dataaslinya takut kalah ya	negative
2	ayo semangat santri	positive
3	repot ya emang	negative
4	dengar santri ktp ngaji cari barokah	negative
5	hataman kak	positive
6	kendit santri turun	negative
7	guna pakai voice	negative
8	jaga ketat bos bermacam-macam pakai	negative
9	ngomong ktp gak banget	negative
10	ktp sampek ngebacot ktp coblos	negative
11	cari amal kak rio	negative
12	kalok gak takut	negative
13	kosong tumbang kosong ayok masyarakat situbondo kuat lanjut periode	negative
14	juang ya	netral
15	kepo bos urus	negative
16	ayo semangat santri blusukan situbondo maju yok kosong	positive
17	terus juang ulama salah paham iya santri masuk paksa adukung calon	positive
18	santri serbaguna nama suruh aja	positive
19	lapor biar bukti	negative

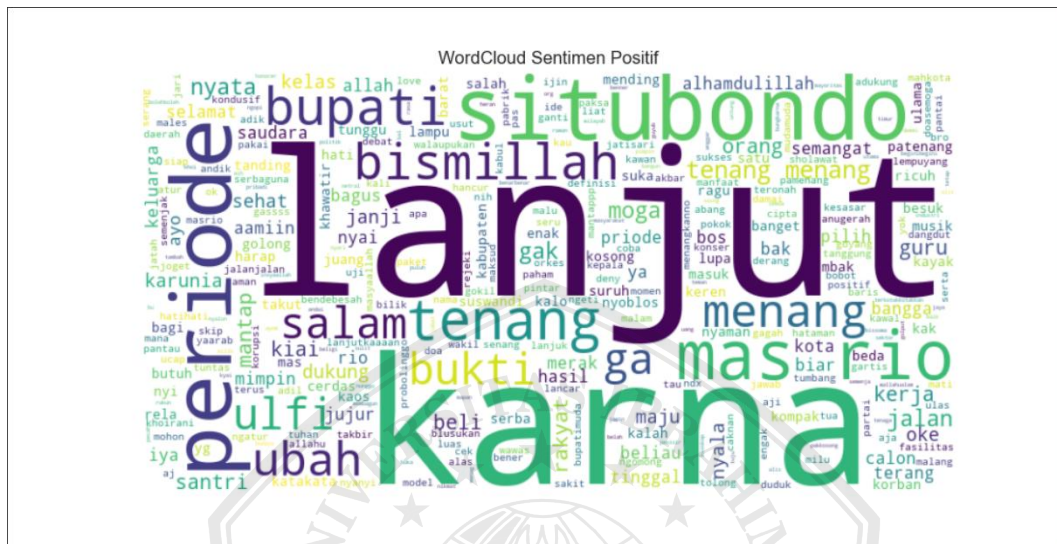
Gambar 4. 5 Hasil Labeling Data

Berikut merupakan distribusi data berdasarkan hasil penandaan manual secara visual ditampilkan pada Gambar 4.6.

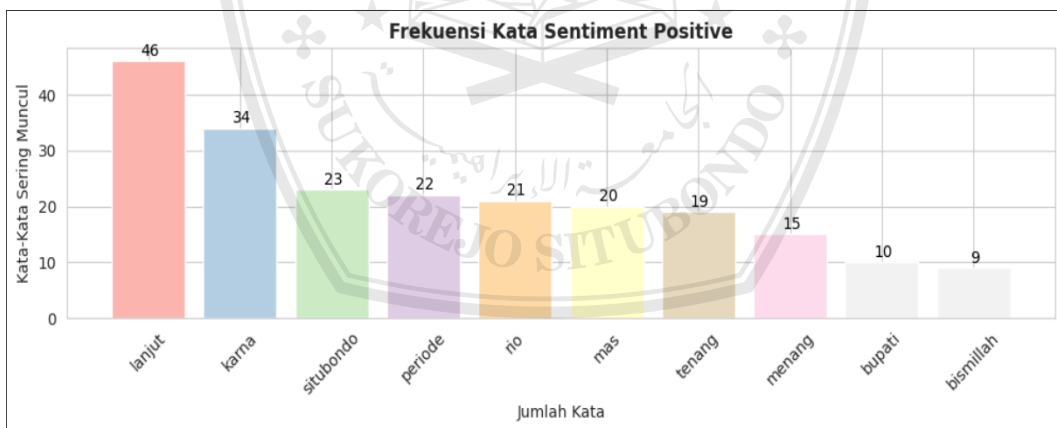


Gambar 4. 6 Diagram Batang Klasifikasi Sentimen

Berikut merupakan hasil *wordcloud* dari klasifikasi sentimen positif ditunjukkan pada gambar 4.7, dan pada gambar 4.8 merupakan frekuensi kata dari sentimen positif.

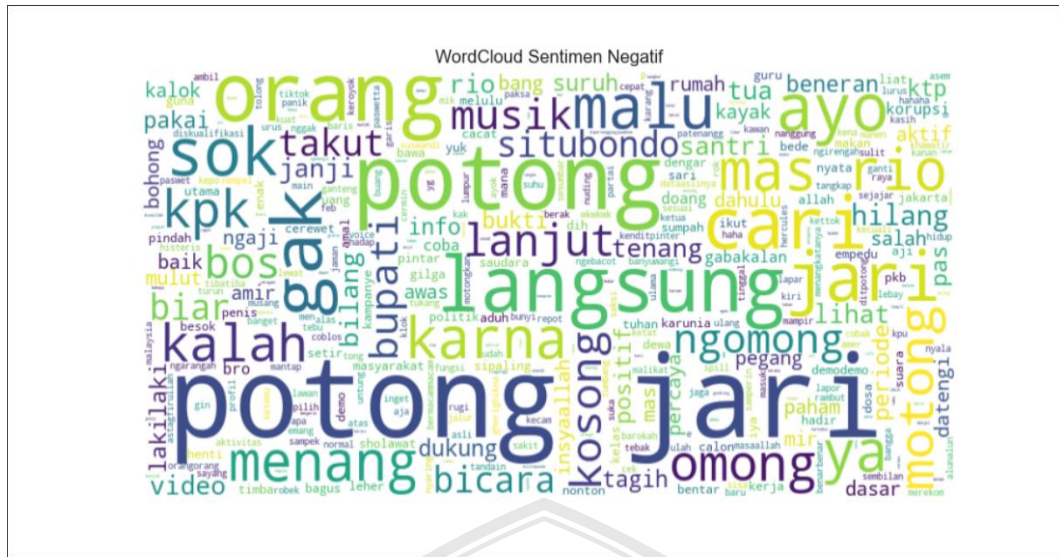


Gambar 4.7 *Wordcloud* Sentimen Positif

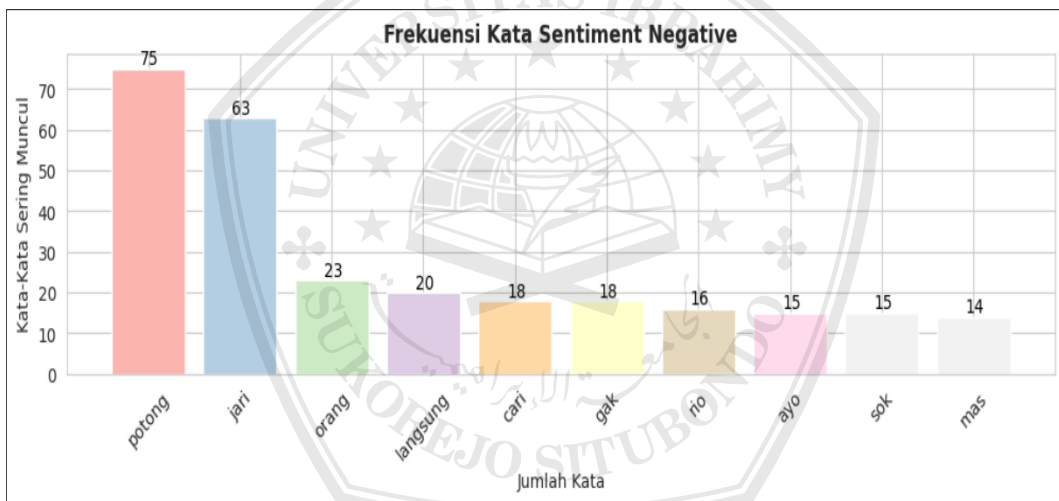


Gambar 4.8 Frekuensi Kata Sentimen Positif

Berikut merupakan hasil *wordcloud* dari klasifikasi sentimen negatif ditunjukkan pada gambar 4.9, dan pada gambar 4.10 merupakan frekuensi kata dari sentimen negatif.

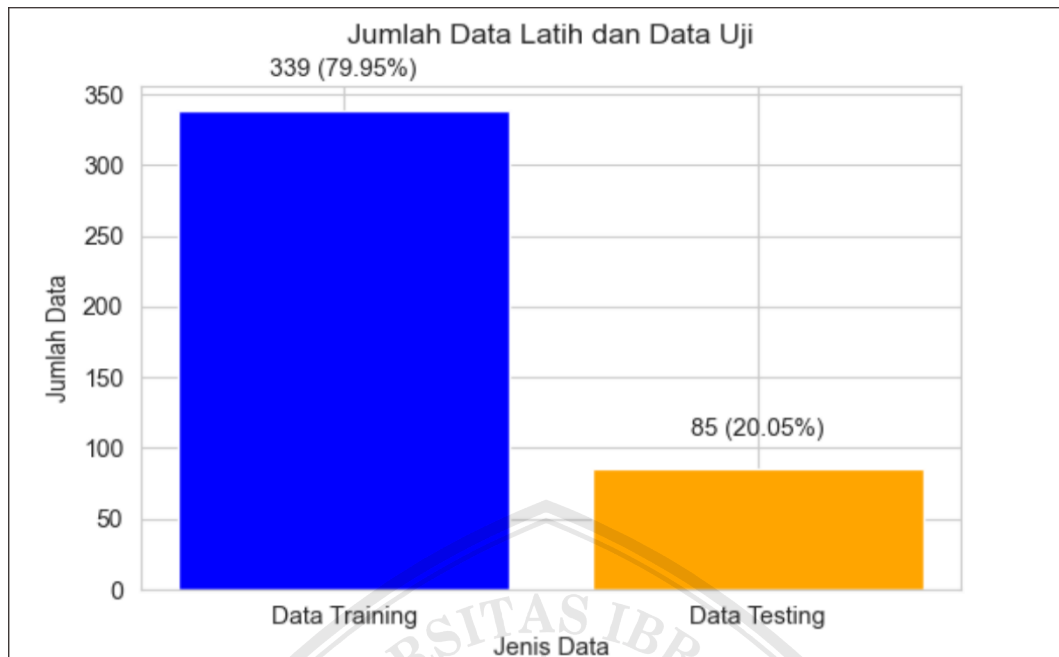


Gambar 4. 9 Wordcloud Sentimen Negatif



Gambar 4. 10 Frekuensi Kata Sentimen Negatif

Berikut merupakan hasil *wordcloud* dari klasifikasi sentimen netral ditunjukkan pada gambar 4.11, dan pada gambar 4.12 merupakan frekuensi kata dari sentimen netral.



Gambar 4. 13 Diagram Batang Split Data.

Fungsi `train_test_split` digunakan untuk memisahkan dataset menjadi data training dan data testing guna melakukan proses pemisahan data ini. Setelah proses *split* selesai, data sentimen yang berkaitan dengan pendapat *netizen* pada komentar terkait PILKADA di Kabupaten Situbondo di *platform* TikTok dibagi menjadi dua bagian 85 data poin yang tersisa, atau sekitar 20,05%, digunakan sebagai data *testing*, sementara 339 data poin, atau sekitar 79,95%, digunakan sebagai data *training*. Dibawah ini merupakan segmen program 4.6 split data.

Segmen Program 4. 6 Split Data

```
1 from sklearn.model_selection import train_test_split
2 # split data jadi data testing dan training
3 x_train, x_test, y_train, y_test =
4 train_test_split(df['stemming data'],
5                 df['Sentimen'],
6                 test_size=0.2, random_state=42)
7 #menyimpan data training
8 train_set = pd.DataFrame({'text': x_train, 'Sentimen':
9 y_train})
10 train_set.to_csv('train_set.csv', index=False)
```

Segmen Program 4.6 (Lanjutan)

```
11 # Simpan data testing
12 test_set = pd.DataFrame({'text': x_test, 'Sentimen':
13 y_test})
14 test_set.to_csv('test_set.csv', index=False)
15 # Menampilkan informasi jumlah data
16 print(f'Jumlah Data Latih: {len(x_train)}')
17 print(f'Jumlah Data Uji: {len(x_test)}')
18 import matplotlib.pyplot as plt
19 # Jumlah data latih dan data uji
20 train_size = len(x_train)
21 test_size = len(x_test)
22 # Membuat plot
23 plt.figure(figsize=(6, 4))
24 bars = plt.bar(['Data Training', 'Data Testing'],
25 [train_size, test_size], color=['blue', 'orange'])
26 # Menambahkan label untuk setiap bar (dalam kurung
27 persentase)
28 for bar in bars:
29     height = bar.get_height()
30     plt.text(bar.get_x() + bar.get_width()/2, height + 20,
31             f'{height} ({height / (train_size + test_size)
32 * 100:.2f})%',
33             ha='center', va='bottom')
34 plt.title('Jumlah Data Latih dan Data Uji\n')
35 plt.xlabel('Jenis Data')
36 plt.ylabel('Jumlah Data')
37 plt.show()
```

4.6 Metode Naive Bayes Classifier

Setelah data komentar melalui proses *preprocessing*, pelabelan data, dan split data menjadi data training dan data testing, langkah selanjutnya adalah menerapkan algoritma *Naive Bayes*. Dimulai dengan mengubah data teks menjadi bentuk numerik dengan menghitung frekuensi kemunculan setiap token (kata unik) dalam dokumen. Hasilnya adalah sebuah matriks frekuensi yang siap digunakan sebagai input model.

Langkah selanjutnya adalah menghitung probabilitas awal (prior) dari masing-masing kelas sentimen. Probabilitas prior dihitung berdasarkan proporsi

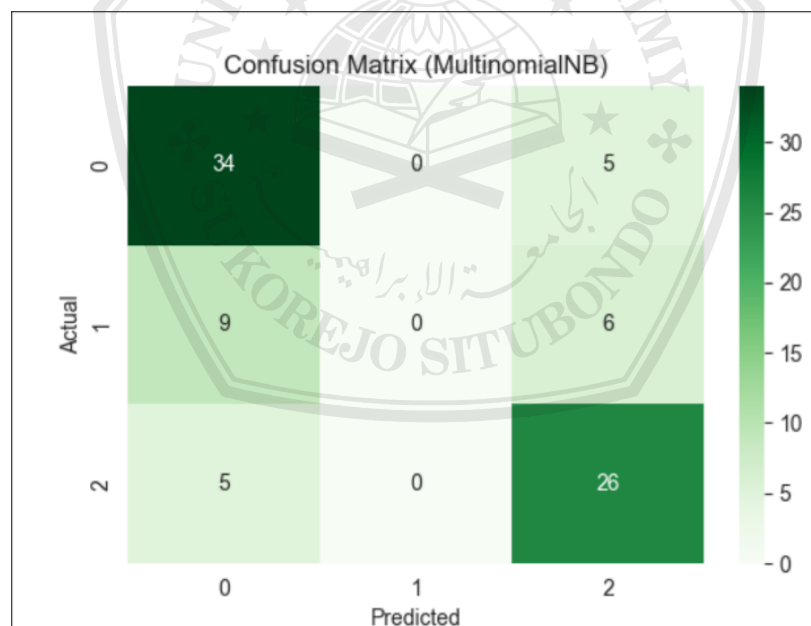
jumlah dokumen yang termasuk dalam suatu kelas terhadap total dokumen pada data *training* (3.1). Setelah itu, dilakukan perhitungan probabilitas kondisional (*likelihood*) dari setiap token terhadap kelas tertentu dengan menggunakan *Laplace smoothing*. Teknik ini digunakan untuk menghindari nilai nol pada token yang tidak muncul di kelas tertentu (3.2).

Dengan menggunakan *TfidfVectorizer* untuk mengubah data teks yang telah melalui beberapa proses di ubah menjadi representasi numerik menggunakan pendekatan TF-IDF (*Term Frequency-Inverse Document Frequency*). Proses ini menghasilkan matriks fitur yang menggambarkan bobot dan signifikansi setiap kata dalam dokumen. Selain itu, teknik *train_test_split* digunakan untuk membagi data menjadi 80% data *training* dan 20% data *testing*. Model klasifikasi yang digunakan adalah metode Multinomial Naive Bayes, yang diinisialisasi menggunakan *MultinomialNB()* dan dilatih menggunakan data *training*. Setelah model dibuat, prosedur prediksi diterapkan pada data *testing* untuk melihat hasil klasifikasi yang dihasilkan dari pelatihan model. Dibawah ini Gambar 4.14 merupakan hasil dari prediksi klasifikasi Naive Bayes menggunakan MultinomialNB.

	stemming data	Actual	Predicted
154	normal bilang potong jari	negative	negative
294	aduh berat kota santri isi konser musik	negative	positive
184	mantap mas karna ok	positive	positive
392	masyarakat gak lihat	negative	negative
439	orang banyuwangi kerja orang	netral	negative
78	lupa lanjut	positive	positive
140	info jari bos	negative	negative
145	barusan	netral	negative
30	kayak enak	positive	negative
77	potong jari	negative	negative
75	info rumah sakit motong jari wkwkwk	negative	negative
100	bukti potong jari mas rio bupati	negative	negative
330	lanjut	positive	positive
96	jaman patenangg insyaallah menang saksi potong jari awas ya	negative	negative
395	ubah	positive	positive
435	beli kaos tenang mas rio guru	positive	positive
9	ngomong ktp gak banget	negative	negative
261	sok	negative	negative
209	lihat salah potong jari	negative	negative
244	dih sok ayo motong jari	negative	negative

Gambar 4. 14 Hasil Prediksi Pada Data Sentimen

Tahap berikutnya adalah mengevaluasi kinerja model *Multinomial Naive Bayes* setelah dilatih dan menghasilkan prediksi. Untuk menentukan distribusi prediksi akurat dan tidak akurat di antara kelas-kelas, evaluasi dilakukan dengan menghitung *confusion matrix* Gambar 4.15. Selain itu, metrik evaluasi seperti akurasi, recall, dan f1-score diperoleh dari setiap kelas menggunakan fungsi *classification_report*. Dengan memanfaatkan fungsi *accuracy_score* digunakan untuk menentukan akurasi model dan menampilkan persentase total prediksi model yang sesuai dengan label aktual ditunjukkan pada gambar 4.16. Hasil evaluasi memberikan ringkasan yang menyeluruh tentang seberapa baik performa model dalam mengklasifikasikan data *testing*.



Gambar 4. 15 Confusion Matrix

```

Classification Report (MultinomialNB):
              precision    recall  f1-score   support

 negative     0.71      0.87      0.78      39
   netral     0.00      0.00      0.00      15
 positive     0.70      0.84      0.76      31

 accuracy                   0.71      85
 macro avg     0.47      0.57      0.52      85
 weighted avg     0.58      0.71      0.64      85

=====
Accuracy (MultinomialNB): 0.7059
=====
    
```

Gambar 4. 16 Classification Report

Tahap selanjutnya merupakan perhitungan manual untuk menentukan akurasi model klasifikasi sentimen. Jumlah prediksi yang akurat dibandingkan dengan jumlah total data testing untuk menentukan akurasi. Pada hasil confusion matrix menunjukkan jumlah prediksi yang benar negatif 34, netral 0, dan positif 26 dengan total 60 data yang dikategorikan dengan akurat. Ada 85 data uji secara keseluruhan. Oleh karena itu, rumus (2.2) digunakan untuk menentukan akurasi model, menghasilkan nilai akurasi 0,7059 atau 70,59%. Angka ini menunjukkan bahwa akurasi klasifikasi data sentimen model secara keseluruhan adalah 71%, yang baik untuk data tiga kelas.

$$Akurasi = \frac{34 + 0 + 26}{34 + 0 + 5 + 9 + 5 + 0 + 0 + 6 + 26} \times 100\% =$$

$$Akurasi = \frac{60}{85} \times 100\% =$$

$$Akurasi = \frac{60}{85} \times 100\% = 0,7059 \text{ atau } 70,59\%$$

Berikut merupakan perhitungan manual *classification report* seperti *precision*, *recall*, dan *f1-score* pada tiap kelas :

a. Negatif

Pada hasil *confusion matrix* negatif menunjukkan jumlah TP untuk negatif=34, sedangkan FP untuk negatif 9+5=14. Oleh karena itu, rumus (2.3) digunakan untuk menentukan *precision* model, menghasilkan nilai *precision* 0.71 atau sekitar 70.83%.

$$Precision = \frac{34}{34 + 14} = 0,7083$$

Selanjutnya menentukan *recall* dengan FN untuk negatif =5. Oleh karena itu, rumus (2.4) digunakan untuk menentukan *recall* model, menghasilkan nilai *recall* 0.87 atau 87.18%.

$$Recall = \frac{34}{34 + 5} = 0.8718$$

Selanjutnya menentukan *F1-score* dengan menghitung hasil *precision* dan *recall* rumus (2.4) digunakan untuk menentukan *F1-score* model, menghasilkan nilai *F1-score* 0.7816 atau 78.16%.

$$F1 = 2x \frac{0.7083 x 0.8718}{0.7083 + 0.8718}$$

$$F1 = 2x \frac{0.6176}{1.5801} = 2x0.3908 = 0.7816$$

b. Netral

Pada hasil *confusion matrix* netral menunjukkan jumlah TP untuk netral=0, sedangkan FP untuk netral =0. Oleh karena itu, rumus (2.3) digunakan untuk menentukan *precision* model, menghasilkan nilai *precision* 0.00 atau sekitar 0.00%.

$$Precision = \frac{0}{0 + 0} = 0.00$$

Selanjutnya menentukan *recall* dengan FN untuk negatif $9+6=15$. Oleh karena itu, rumus (2.4) digunakan untuk menentukan *recall* model, menghasilkan nilai *recall* 0.00 atau 0.00%.

$$Recall = \frac{0}{0 + 15} = 0.00$$

Selanjutnya menentukan *F1-score* dengan menghitung hasil *precision* dan *recall* rumus (2.4) digunakan untuk menentukan *F1-score* model, menghasilkan nilai *F1-score* 0.00 atau 0.00%

$$F1 = 2x \frac{0}{0} = 2x0 = 0.00$$

c. Positif

Pada hasil *confusion matrix* negatif menunjukkan jumlah TP untuk positif=26, sedangkan FP untuk positif $6+5=11$. Oleh karena itu, rumus (2.3) digunakan untuk menentukan *precision* model, menghasilkan nilai *precision* 0.70 atau sekitar 70.27%.

$$Precision = \frac{26}{26 + 11} = 0,7027$$

Selanjutnya menentukan *recall* dengan FN untuk negatif =5. Oleh karena itu, rumus (2.4) digunakan untuk menentukan *recall* model, menghasilkan nilai *recall* 0.84 atau 83.87%.

$$Recall = \frac{26}{26 + 5} = 0.8387$$

Selanjutnya menentukan *F1-score* dengan menghitung hasil *precision* dan *recall* rumus (2.4) digunakan untuk menentukan *F1-score* model, menghasilkan nilai *F1-score* 0.7647 atau 76.47 %

$$F1 = 2x \frac{0.7027x 0.8387}{0.7027 + 0.8387}$$

$$F1 = 2x \frac{0.5894}{1.5414} = 2x0.3824 = 0.7647$$

Pada segmen program 4.7 dibawah ini merupakan kode untuk menerapkan algoritma *Naive Bayes MultinomialNB*.

Segmen Program 4. 7 *Naive Bayes*

```
1 import pandas as pd
2 import numpy as np
3 from sklearn.model_selection import train_test_split
4 from sklearn.naive_bayes import MultinomialNB
5 from sklearn.metrics import confusion_matrix,
6 classification_report, accuracy_score
7 from sklearn.feature_extraction.text import TfidfVectorizer
8 import seaborn as sns
9 import matplotlib.pyplot as plt
10 from tabulate import tabulate
11
12 # Preprocessing data
13 tfidf = TfidfVectorizer()
14 x = tfidf.fit_transform(df['stemming data']).toarray()
15 y = df['Sentimen']
16
17 # Split data
18 X_train, X_test, y_train, y_test = train_test_split(x, y,
19 test_size=0.2, random_state=42)
20
21 # Initialize and train model MultinomialNB
22 mnb = MultinomialNB()
23 mnb.fit(X_train, y_train)
24
25 # Predict with MultinomialNB
26 y_pred_mnb = mnb.predict(X_test)
27 # Evaluate MultinomialNB
28 conf_matrix_mnb = confusion_matrix(y_test, y_pred_mnb)
29 class_report_mnb = classification_report(y_test, y_pred_mnb)
30 accuracy_mnb = accuracy_score(y_test, y_pred_mnb)
31
32 print("MultinomialNB Results")
33 print("=====")
```

Segmen Program 4.13 (Lanjutan)

```
34 print("Confusion Matrix (MultinomialNB):")
35 print(conf_matrix_mnb)
36 print("=====")
37 print("\nClassification Report (MultinomialNB):")
38 print(class_report_mnb)
39 print("=====")
40 print(f"Accuracy (MultinomialNB): {accuracy_mnb:.4f}")
41 print("=====")
42
43 # Plot confusion matrix for MultinomialNB
44 plt.figure(figsize=(6, 4))
45 sns.heatmap(conf_matrix_mnb, annot=True, fmt='d',
46             cmap='Greens')
47 plt.title('Confusion Matrix (MultinomialNB)')
48 plt.xlabel('Predicted')
49 plt.ylabel('Actual')
50 plt.show()
51
52 # Create DataFrame for actual vs predicted (MultinomialNB)
53 results_mnb = pd.DataFrame({
54     'stemming data': df.loc[y_test.index, 'stemming data'],
55     'Actual': y_test,
56     'Predicted': y_pred_mnb
57 })
58 results_mnb.to_csv('Hasil_pred_MultinomialNB.csv',
59                  encoding='utf8', index=False)
60 print("Actual vs Predicted (MultinomialNB):")
61 print(tabulate(results_mnb.head(20), headers='keys',
62               tablefmt='github'))
```

BAB V

PENUTUP

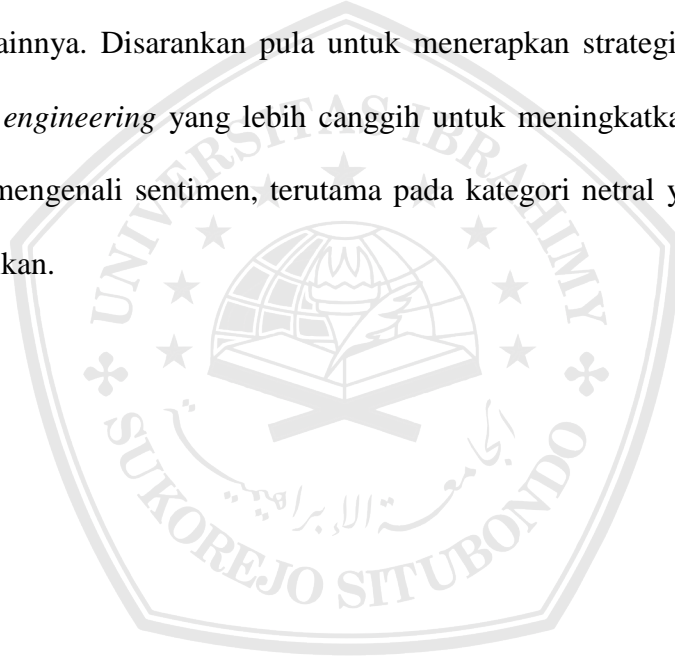
5.1 Kesimpulan

Metode *Naive Bayes Classifier* digunakan secara efektif dalam penelitian ini untuk melakukan analisis sentimen pada PILKADA di Kabupaten Situbondo tahun 2024 di *platform TikTok*. Data yang diperoleh melalui proses *scraping* menggunakan *Data Miner* berjumlah 494 komentar, yang kemudian data di *Preprocessing* meliputi *Cleansing, Case Folding, Tokenizing, Remove Stopword, Stemming Data* dan dikategorikan ke dalam tiga kategori sentimen *positive, negative, dan neutral*. Teknik *Multinomial Naïve Bayes* digunakan untuk membuat model, dengan *split data* menjadi data *training* sebesar 79,95% dan *testing* 20,05%. Berdasarkan hasil evaluasi menghasilkan akurasi model mencapai 70,59%, dan menghasilkan nilai presisi, *recall*, dan *F1-score* yang berbeda-beda tiap kategori sentimen.

Hasil penelitian menunjukkan bahwa teknik *Naïve Bayes* cukup efektif dalam mengidentifikasi sentimen komentar politik di media sosial, dan masih ada potensi untuk peningkatan nilai akurasi, terutama dalam kategori sentimen netral yang kurang terklasifikasikan dengan baik. Sehingga para kandidat, staf kampanye, dan penyelenggara PILKADA dapat memanfaatkan temuan penelitian ini untuk mengidentifikasi isu-isu yang muncul, merancang strategi kampanye dalam segi komunikasi politik yang lebih tepat sasaran, dan memperoleh pemahaman yang lebih terukur tentang opini publik.

5.2 Saran

Saran pada penelitian selanjutnya agar bisa mengembangkan model analisis sentimen dengan akurasi yang lebih tinggi. Tidak hanya menggunakan metode *Naive Bayes*, diharapkan bisa membandingkan berbagai teknik klasifikasi lainnya termasuk *Random Forest*, *Support Vector Machine (SVM)*, atau algoritma berbasis *deep learning*. Untuk membuat hasil penelitian lebih mencerminkan opini publik, juga dimungkinkan untuk memperluas jumlah dan variasi data dari *platform* media sosial lainnya. Disarankan pula untuk menerapkan strategi optimasi atau teknik *feature engineering* yang lebih canggih untuk meningkatkan kemampuan model dalam mengenali sentimen, terutama pada kategori netral yang seringkali sulit dikategorikan.

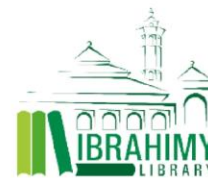


Daftar Pustaka

- [1] S. Suyatno, "Pemilihan Kepala Daerah (Pilkada) dan Tantangan Demokrasi Lokal di Indonesia," *Polit. Indones. Indones. Polit. Sci. Rev.*, vol. 1, no. 2, p. 212, 2016, doi: 10.15294/jpi.v1i2.6586.
- [2] N. Apriliani, N. Suarna, and W. Prihartono, "Analisis Sentimen Review Penggunaan Tiktok Melalui Pendekatan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 6, pp. 3725–3731, 2024, doi: 10.36040/jati.v7i6.8299.*
- [3] M. S. P. Yoga et al., "PEMANFAATAN TIKTOK," vol. 6, no. 3, pp. 57–77, 2021.
- [4] A. Khusnul Khotimah, "Analisis Sentimen Terhadap Kualitas Pelayanan," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 3, pp. 3044–3048, 2024, doi: 10.36040/jati.v8i3.9520.*
- [5] S. Kusuma Wardani and Y. Arum Sari, "Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5582–5590, 2021.
- [6] M. H. K. S. Fadila, Woro Isti Rahayu, Penerapan Metode Naive Bayes dan Skala Likert Pada Aplikasi Prediksi Kelulusan Mahasiswa. Indonesia: Kreatif Industri Nusantara.
- [7] M. C. Auliya Rahman Isnain, "Pemanfaatan Text Mining pada Pemilu 2024: Mengoptimalkan Analisis Data untuk Pemilihan yang Lebih Cerdas," *FTIK Teknokrat.*
- [8] S. H. Eni Puji Estuti, Wachidah Fauziyanti, Analisis Deskriptif dan Kuantitatif Produktivitas Garam Indonesia : Studi Kasus pada Petani Garam Kabupaten Pati. Penerbit NEM, 2021.
- [9] M. F. Arib, M. S. Rahayu, R. A. Sidorj, and M. W. Afgani, "Experimental Research Dalam Penelitian Pendidikan," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 1, pp. 5497–5511, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/8468>
- [10] S. J. Gellysa Urva, Desyanti, Isa Albanna, Muchamad Sobri Sungkar, I Made Agus Oka Gunawan, Iwan Adhichandra, Sahrul Ramadhan, Rifky Lana Rahardian, Herlawati, Rahmadya Trias Handayanto, Anak Agung Gede Bagus Ariana, Hartatik, Prima Dina Atika, PENERAPAN DATA MINING DI BERBAGAI BIDANG : Konsep, Metode, dan Studi Kasus. PT. Sonpedia Publishing Indonesia, 2023.

- [11] N. Firdha, Damira, R. Fitri, G. H. Selaras, and I. G. N. Saputra, "Studi Literatur Tentang Peningkatan Kompetensi Belajar Peserta Didik Melalui Kegiatan Pembelajaran Kolaboratif Berbasis Lesson Study," *Pros. SEMNAS BIO*, vol. 01, pp. 1005–1013, 2021.
- [12] A. N. Izzati, "Analisis Sentimen Hasil Putusan Mk Terkait Sengketa Pilpres 2024 Pada Media Sosial X," *J. Teknol. Inf. Indones.*, vol. 9, no. 1, pp. 43–50, 2024, doi: 10.30869/jtii.v9i1.1338.
- [13] C. L. Rohmat, R. Tito Aprilianto, F. Fathurrohman, and I. Iin, "Analisis Sentimen Pada Ulasan Produk Pakaian Zirdigo Dengan Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 668–674, 2024, doi: 10.36040/jati.v8i1.8697.
- [14] A. Nurian, M. S. Ma'arif, I. N. Amalia, and C. Rozikin, "Analisis Sentimen Pengguna Aplikasi Shopee Pada Situs Google Play Menggunakan Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3631.
- [15] M. T. Muhammad Arhami, S.Si., M.Kom., Muhammad Nasir, S.T., *Data Mining - Algoritma dan Implementasi*. Indonesia: Andi Offset.
- [16] R. M. Sari, *Klasifikasi Data Mining*. Indonesia: Serasi Media Teknologi, 2024. [Online]. Available: https://www.google.co.id/books/edition/Klasifikasi_Data_Mining/xTwXEQAAQBAJ?hl=id&gbpv=0
- [17] dkk Yessy Asri, S.T., MMSI, Dr. Dra. Dwina Kuswardani, M.Kom, *MACHINE LEARNING & DEEP LEARNING: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Indonesia: Uwais Inspirasi Indonesia, 2024. [Online]. Available: https://www.google.co.id/books/edition/MACHINE_LEARNING_DEEP_LEARNING_Analisis/Yu7uEAAAQBAJ?hl=id&gbpv=0&kptab=overview
- [18] C. P. Ema Ainun Novia, Woro Isti Rahayu, *SISTEM PERBANDINGAN ALGORITMA K-MEANS DAN NAÏVE BAYES UNTUK MEMPREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN*. Kreatif Industri Nusantara, 2020.
- [19] T. Arifqi, N. Suarna, and W. Prihartono, "Penggunaan Naive Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Mcdonald'S Di Indonesia," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1949–1956, 2024, doi: 10.36040/jati.v8i2.8740.
- [20] E. Junaidi Surya, *DASAR-DASAR PEMROGRAMAN DENGAN PYTHON*. PT. Sonpedia Publishing Indonesia, 2023.

- [21] Q. Budiman, S. Mouton, L. Veenhoff, and A. Boersma, “程威特 1 , 吴海涛 1 , 江帆 2,” J. Inov. Penelit., vol. 1, no. 0.1101/2021.02.25.432866, pp. 1–15, 2021.
- [22] P. P. Endang Supriyadi, Panduan Praktis Belajar Microsoft Excel 2021. Indonesia: TOHAR MEDIA.
- [23] M. S. Pardianti and V. V. S, “Pengelolaan Konten Tiktok Sebagai Media Informasi,” Ikon --Jurnal Ilm. Ilmu Komun., vol. 27, no. 2, pp. 187–210, 2022, doi: 10.37817/ikon.v27i2.1905.
- [24] S. M. K. Yahya, Data Mining. CV Jejak (Jejak Publisher).
- [25] dan Z. Ben, David, Ken, “Data Miner adalah alat pengikisan web paling kuat untuk penambang data profesional,” Innovation Lab LLC -Seattle.
- [26] M. B. Nendya, B. Susanto, G. I. W. Tamtama, and T. J. Wijaya, “Desain Level Berbasis Storyboard Pada Perancangan Game Edukasi Augmented Reality Tap The Trash,” Fountain Informatics J., vol. 8, no. 1, pp. 1–6, 2023, doi: 10.21111/fij.v8i1.8836.
- [27] F. Alghifari and D. Juardi, “Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes,” J. Ilm. Inform., vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [28] A. Cahya Kamilla, N. Priyani, R. Priskila, and V. Handrianus Pranatawijaya, “Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naïve Bayes Classifier,” JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 3, pp. 2923–2928, 2024, doi: 10.36040/jati.v8i3.9587.
- [29] A. S. dan N. Surojudin, “Analisis Dan Perbandingan Stemming Algoritma Porter Dengan Algoritma Ahmad Yusoff Sembok Dalam Dokumen Teks Bahasa Indonesia,” Pros. Semin. SeNTIK, vol. 4, no. 1, pp. 347–357, 2020, [Online]. Available: <https://ejournal.jakstik.ac.id/index.php/sentik/article/view/3304>
- [30] Suyanto, Machine learning : tingkat dasar dan lanjut. INFORMATIKA, 2018.

**SURAT KETERANGAN
HASIL PEMERIKSAAN PLAGIASI**

Yang bertanda tangan di bawah ini

Nama : Muhammad Ali Ridla, M.Kom.

Jabatan : Kepala Perpustakaan

Menyatakan dengan sebenarnya bahwa:

NIM : 2021503113

Nama : MA'RUF UBAIDILLAH

Fakultas : Sains dan Teknologi

Prodi : Teknologi Informasi

Kecamatan : JANGKAR

Kabupaten : KAB. SITUBONDO

Provinsi : Jawa Timur

Judul Skripsi : ANALISIS SENTIMEN MENGGUNAKAN
ALGORITMA NAIVE BAYES PADA PILKADA
KABUPATEN SITUBONDO TAHUN

Dengan dosen Pembimbing :

1. Ahmad Baijuri, M.Kom.

2. Nur Azise, M.Kom.

Telah dilakukan cek plagiasi di Perpustakaan Universitas Ibrahimi dengan persentase plagiasi terakhir sebesar **23%** .

Demikian Surat Keterangan ini dibuat untuk dipergunakan sebagaimana mestinya.

Sukorejo, 16 Agustus 2025

Kepala Perpustakaan,



Muhammad Ali Ridla, M.Kom.