

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Sentiment analysis, atau opinion mining, merupakan proses otomatis untuk mengidentifikasi dan memahami sentimen dalam data teks, baik positif maupun negatif. Dengan perkembangan teknologi, masyarakat kini lebih mudah menyampaikan pendapat melalui media sosial seperti X, yang memungkinkan pengguna berbagi pesan teks hingga 280 karakter. Tweet yang dibuat masyarakat sering kali membahas berbagai topik, termasuk opini tentang Ibu Kota Nusantara. Mengidentifikasi pendapat masyarakat tentang Ibu Kota Nusantara menjadi penting untuk memahami opini publik terkait pembangunan dan perkembangan ibu kota baru ini. Topik yang dibahas adalah berbagai aspek yang berkaitan dengan Ibu Kota Nusantara, seperti infrastruktur, kebijakan pemerintah, dan dampak sosial. Sementara itu, sentimen adalah nilai dari *tweet* tersebut yang menunjukkan apakah *tweet* itu bersifat positif atau negatif. Dengan demikian, analisis sentimen mampu memberikan pemahaman yang lebih mendalam mengenai bagaimana masyarakat merespon isu-isu terkait Ibu Kota Nusantara, yang dapat membantu dalam pengambilan keputusan dan strategi komunikasi yang lebih baik.

3.2 Hasil Analisis

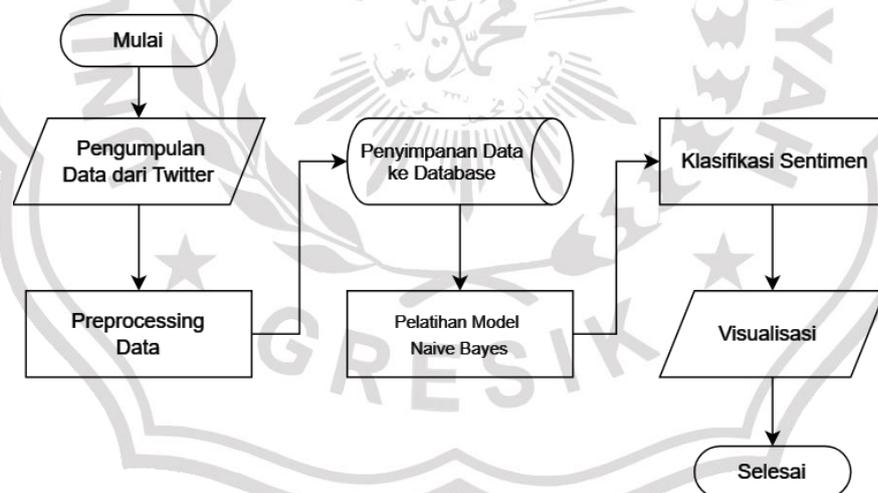
Hasil Analisis sentimen terhadap tweet yang membahas Ibu Kota Nusantara bertujuan untuk mengklasifikasikan sentimen yang terkandung dalam tweet di platform media sosial X. Sistem klasifikasi ini dikembangkan menggunakan metode Naïve Bayes, yang memanfaatkan data tweet berlabel sentimen sebagai data latih atau acuan untuk menentukan kelas sentimen pada tweet baru. Dengan metode ini, setiap tweet akan dikategorikan ke beberapa sentimen, contoh positif, netral, atau negatif. Klasifikasi tersebut dapat digunakan untuk menganalisis opini publik mengenai pembangunan dan perkembangan Ibu Kota Nusantara.

Pengumpulan informasi dilakukan melalui observasi langsung terhadap tanggapan pengguna media sosial X terkait pemindahan Ibu Kota Nusantara (IKN). Proses observasi berlangsung dari 1 Januari 2024 hingga 17 Agustus 2024, dengan total 4.180 komentar dari masyarakat yang dianalisis.

3.3 Perancangan Sistem

3.3.1 Flowchart Sistem

Sistem yang dikembangkan adalah sebuah aplikasi untuk menganalisis sentimen terkait Ibu Kota Nusantara di media sosial X, dengan menggunakan metode Naive Bayes Classifier untuk mengklasifikasikan tweet. Tujuan dari analisis sistem ini adalah merancang aplikasi yang dapat mengevaluasi sentimen publik terhadap Ibu Kota Nusantara (IKN) melalui platform media sosial X. Diharapkan aplikasi ini dapat memberikan wawasan kepada pemerintah dan pihak-pihak terkait mengenai persepsi publik, sehingga dapat mendukung proses pengambilan keputusan yang lebih efektif. Gambaran umum dari sistem yang akan dikembangkan dapat dilihat pada gambar 3.1.

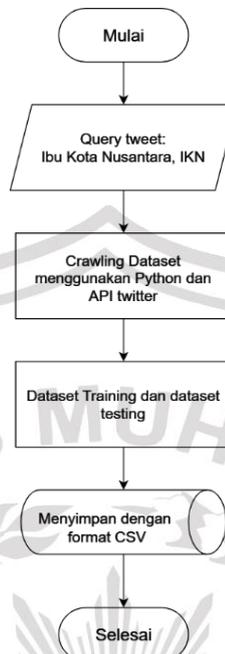


Gambar 3.1 Flowchart Sistem

3.3.2 Representasi Dataset

Pada tahapan ini dilakukannya pengumpulan data dengan cara Crawling menggunakan Python melalui komponen *API Search X*. *Tweet* yang diambil berisi tentang opini mengenai pemindahan ibu kota dari Jakarta ke Kalimantan dengan query “ibu kota nusantara” dan “IKN.” Data Crawling dimulai dari tanggal 01

januari 2024 sampai dengan 17 Agustus 2024 dan mendapatkan sebanyak 4180 data *tweets*.



Gambar 3.2 Alur Crawling Dataset

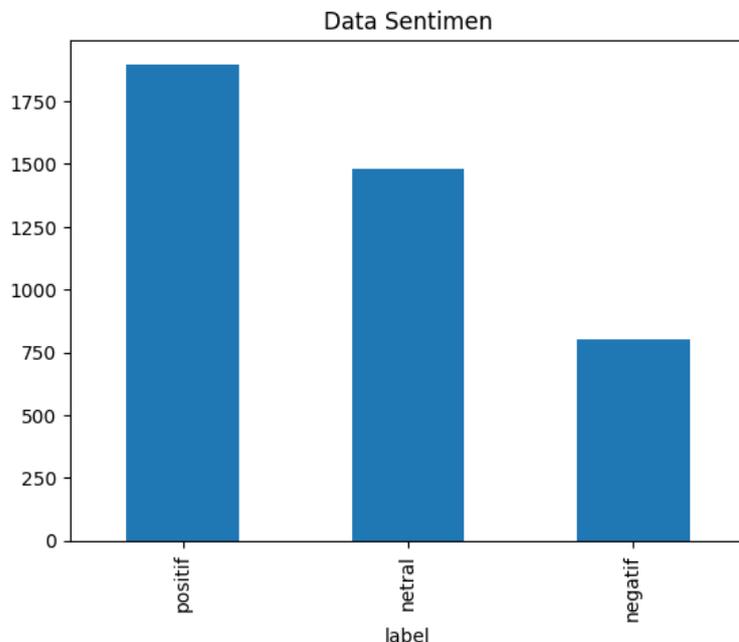
- Langkah pertama adalah menghubungkan ke X dengan menggunakan *API X*. Bagian ini digunakan untuk menyimpan token autentikasi X yang diperlukan agar dapat mengakses *API X*. Token ini memungkinkan aplikasi untuk berkomunikasi dengan *X API*.
- Selanjutnya menentukan *query* yang mau diambil
- Menyimpan dataset ke file berformat csv.

Berikut adalah contoh hasil tabel crawling dataset menggunakan python.

Tabel 3.1 Contoh Hasil *Crawling* Menggunakan Python

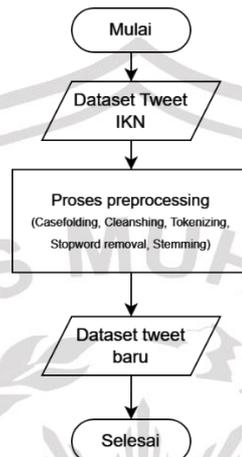
username	full_text
faifrytale	Jelas yaah berarti bilangny si paling pro rakyat tAPI IKN aja gak didukung
AF16001	@ikhwanuddin @innovacomunity Kalo mau bener2 nerAPIn EV ya semua kota harus ada SPKLU.
FarhanAdl	@_yantochan_ @meguru_meruru @UGM_FESS Berarti bisa lebih buruk jg kan? Wong dari programnya Jokowi aja ga ada tuh rencana bikin ikn revisi UU KPK bikin UU cipta kerja

Pada tahap analisis data, setiap cuitan dianalisis untuk mengidentifikasi sentimennya, yang kemudian dikelompokkan ke dalam tiga kategori: positif, netral, dan negatif. Distribusi hasil klasifikasi sentimen tersebut dapat dilihat pada Gambar 3.3.

**Gambar 3.3** Distribusi Data Sentimen terhadap IKN

3.3.3 Preprocessing

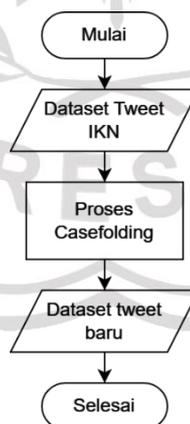
Tahap Preprocessing penting untuk mengurangi atribut yang kurang relevan sebelum klasifikasi. Pada tahap ini, data mentah dibersihkan agar lebih berkualitas, sehingga mempermudah proses klasifikasi. Langkah-langkah preprocessing ditunjukkan pada Gambar 3.4.



Gambar 3.4 Alur *Preprocessing*

3.3.3.1 Casefolding

Pada tahap *casefolding*, semua huruf kapital dalam teks *tweet* diubah menjadi huruf kecil. Diagram alir untuk proses *casefolding* ditunjukkan pada Gambar 3.5.



Gambar 3.5 Alur proses *casefolding*

Sebagai ilustrasi dari proses cleansing, contoh *tweet* yang dihasilkan ditampilkan pada tabel 3.2.

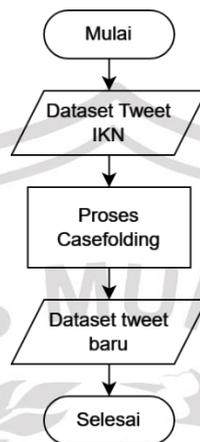
Tabel 3.2 Contoh hasil proses *casefolding*

<i>CaseFolding</i>	
Sebelum	Sesudah
Tugas Pemerintah itu yang utama ada 2 menurut konstitusi Mencerdaskan Kebutuhan Rakyat dan Mensejahterakan Rakyat! Ini Mahasiswa Yg tidak mampu malah disuruh Pinjol !! Biadab !! Bangun IKN Bisa Korupsi BTS Bisa Biayain Food Estate bisa masa biayain anak bangsa kagak mampu!!! https://t.co/c8ChWJC3xj	tugas pemerintah itu yang utama ada 2 menurut konstitusi mencerdaskan kebidupan rakyat dan mensejahterakan rakyat! ini mahasiswa yg tidak mampu malah disuruh pinjol !! biadab !! bangun ikn bisa korupsi bts bisa biayain food estate bisa masa biayain anak bangsa kagak mampu!!! https://t.co/c8chwjc3xj
Jika dia terpilih gue rasa proyek IKN juga kemungkinan akan dibatalkan layaknya pulau reklamasi kemungkinan. karena IKN proyek besar banget dan ambisius pasti akan menantang banget untuk membatalkannya.	jika dia terpilih gue rasa proyek ikn juga kemungkinan akan dibatalkan layaknya pulau reklamasi kemungkinan. karena ikn proyek besar banget dan ambisius pasti akan menantang banget untuk membatalkannya.

3.3.3.2 *Cleansing*

Tahapan Cleansing adalah proses pembersihan data teks dengan menghapus elemen-elemen yang tidak relevan atau tidak diinginkan, seperti tanda baca, simbol, URL, dan karakter khusus lainnya. Penghapusan angka dalam tahap pre-processing untuk analisis sentimen tergantung pada konteks dan tujuan analisis yang dilakukan. Beberapa contoh atribut yang dianggap tidak penting antara lain mention yang dimulai dengan simbol ('@'), hashtag dengan simbol ('#'), URL yang diawali dengan ('http', 'bit.ly'), serta karakter simbol seperti

(~!@#\$\$%^&*())_+?<>.,?:{}[]]). Tujuan dari *Cleansing* adalah untuk memperbaiki kualitas data sehingga mempermudah analisis dan meningkatkan akurasi hasil pada tahap pemrosesan selanjutnya, seperti klasifikasi atau analisis sentimen. Berikut diagram alir *Cleansing* terdapat pada gambar 3.6.



Gambar 3.6 Alur proses *cleansing*

Sebagai ilustrasi dari proses cleansing, contoh *tweet* yang dihasilkan ditampilkan pada tabel 3.3.

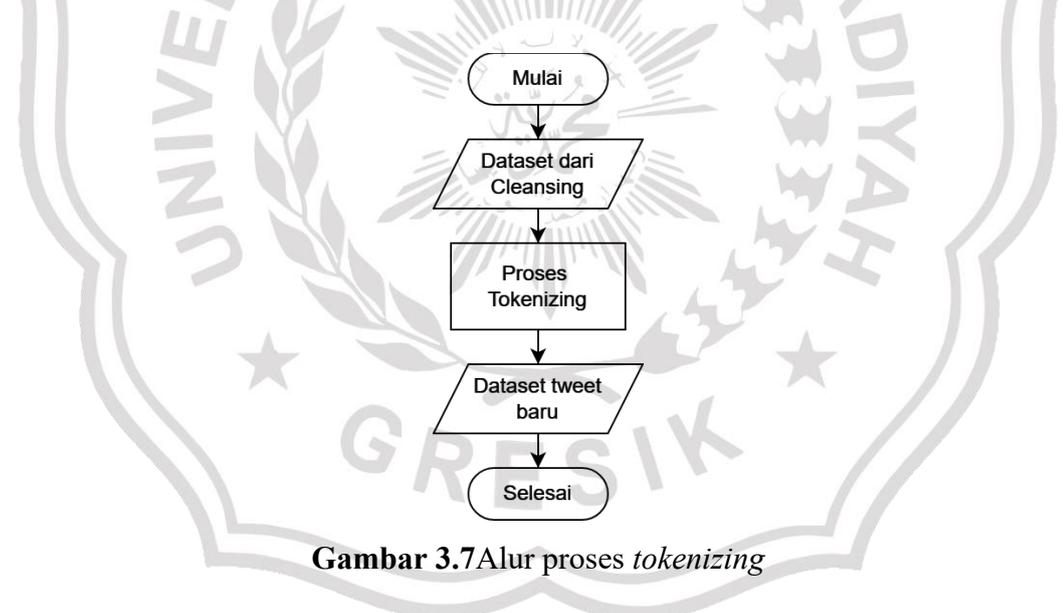
Tabel 3.3 Contoh hasil proses *celansing*

<i>Cleansing</i>	
Sebelum	Sesudah
tugas pemerintah itu yang utama ada 2 menurut konstitusi mencerdaskan kebidupan rakyat dan mensejahterakan rakyat! ini mahasiswa yg tidak mampu malah disuruh pinjol !! biadab !! bangun ikn bisa korupsi bts bisa biayain food estate bisa masa biayain anak bangsa kagak mampu!!! https://t.co/c8chwjc3xj	tugas pemerintah itu yang utama ada 2 menurut konstitusi mencerdaskan kebidupan rakyat dan mensejahterakan rakyat ini mahasiswa yg tidak mampu malah disuruh pinjol biadab bangun ikn bisa korupsi bts bisa biayain food estate bisa masa biayain anak bangsa kagak mampu

jika dia terpilih gue rasa proyek ikn juga kemungkinan akan dibatalkan layaknya pulau reklamasi kemungkinan. karena ikn proyek besar banget dan ambisius pasti akan menantang banget untuk membatalkannya.	jika dia terpilih gue rasa proyek ikn juga kemungkinan akan dibatalkan layaknya pulau reklamasi kemungkinan karena ikn proyek besar banget dan ambisius pasti akan menantang banget untuk membatalkannya
--	--

3.3.3.3 Tokenizing

Proses pemecahan teks menjadi bagian-bagian kecil yang disebut token, yang bisa berupa kata, frasa, atau simbol, tergantung pada kebutuhan analisis. Tokenizing memudahkan pemrosesan teks dengan cara yang lebih terstruktur, sehingga setiap token bisa dianalisis secara terpisah. Proses ini penting untuk berbagai tugas dalam pemrosesan bahasa alami (NLP), seperti analisis sentimen, klasifikasi teks, dan ekstraksi informasi. Diagram alir Tokenizing dapat dilihat pada gambar 3.7.



Gambar 3.7 Alur proses *tokenizing*

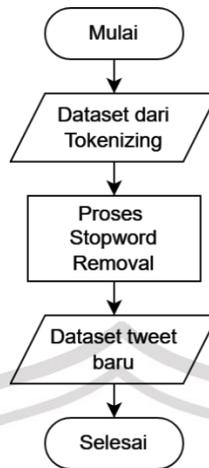
Sebagai ilustrasi dari proses *tokenizing*, contoh *tweet* yang dihasilkan ditampilkan pada tabel 3.4.

Tabel 3.4 Contoh hasil proses *tokenizing*

<i>Tokenizing</i>	
Sebelum	Sesudah
tugas pemerintah itu yang utama ada 2 menurut konstitusi mencerdaskan kebidupan rakyat dan mensejahterakan rakyat ini mahasiswa yg tidak mampu malah disuruh pinjol biadab bangun ikn bisa korupsi bts bisa biayain food estate bisa masa biayain anak bangsa kagak mampu	['tugas', 'pemerintah', 'itu', 'yang', 'utama', 'ada', '2', 'menurut', 'konstitusi', 'mencerdaskan', 'kebidupan', 'rakyat', 'dan', 'mensejahterakan', 'rakyat', 'ini', 'mahasiswa', 'yg', 'tidak', 'mampu', 'malah', 'disuruh', 'pinjol', 'biadab', 'bangun', 'ikn', 'bisa', 'korupsi', 'bts', 'bisa', 'biayain', 'food', 'estate', 'bisa', 'masa', 'biayain', 'anak', 'bangsa', 'kagak', 'mampu']
jika dia terpilih gue rasa proyek ikn juga kemungkinan akan dibatalkan layaknya pulau reklamasi kemungkinan karena ikn proyek besar banget dan ambisius pasti akan menantang banget untuk membatalkannya	['jika', 'dia', 'terpilih', 'gue', 'rasa', 'proyek', 'ikn', 'juga', 'kemungkinan', 'akan', 'dibatalkan', 'layaknya', 'pulau', 'reklamasi', 'kemungkinan', 'karena', 'ikn', 'proyek', 'besar', 'banget', 'dan', 'ambisius', 'pasti', 'akan', 'menantang', 'banget', 'untuk', 'membatalkannya']

3.3.3.4 Stopword Removal

Tahapan Stopword removal adalah proses penghapusan kata-kata umum yang tidak memberikan makna signifikan dalam analisis teks, seperti "dan", "yang", "adalah", dan "di". Meskipun kata-kata ini sering muncul dalam teks, mereka tidak memberikan kontribusi penting terhadap pemahaman atau analisis konteks. Menghapus stopwords dapat meningkatkan efisiensi dan akurasi algoritma pemrosesan teks dengan mengurangi jumlah data yang perlu diproses dan memfokuskan analisis pada kata-kata yang lebih relevan dan informatif. Diagram alir Stopword removal dapat dilihat pada gambar 3.8.



Gambar 3.8 Alur proses *stopword removal*

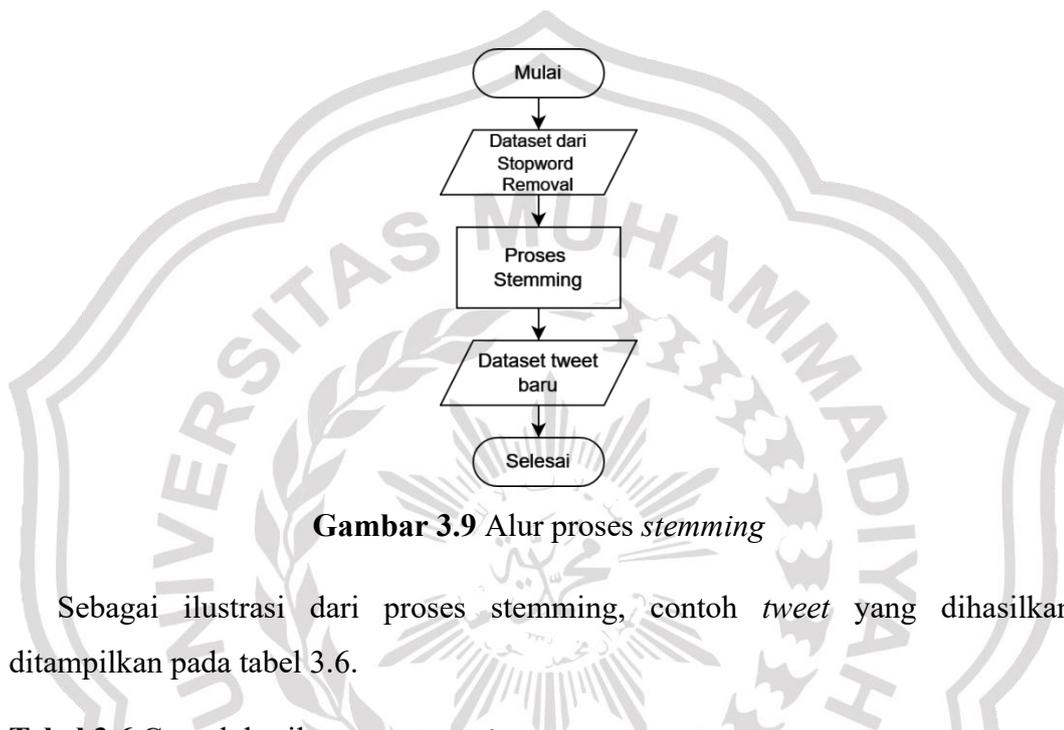
Sebagai ilustrasi dari proses *stopword removal*, contoh *tweet* yang dihasilkan ditampilkan pada tabel 3.5.

Tabel 3.5 Contoh hasil proses *stopword removal*

<i>Stopword Removal</i>	
Sebelum	Sesudah
['tugas', 'pemerintah', 'itu', 'yang', 'utama', 'ada', '2', 'menurut', 'konstitusi', 'mencerdaskan', 'kebidupan', 'rakyat', 'dan', 'mensejahterakan', 'rakyat', 'ini', 'mahasiswa', 'yg', 'tidak', 'mampu', 'malah', 'disuruh', 'pinjol', 'biadab', 'bangun', 'ikn', 'bisa', 'korupsi', 'bts', 'biayain', 'food', 'estate', 'bisa', 'masa', 'biayain', 'anak', 'bangsa', 'kagak', 'mampu']	['tugas', 'pemerintah', 'utama', '2', 'konstitusi', 'mencerdaskan', 'kebidupan', 'rakyat', 'mensejahterakan', 'rakyat', 'mahasiswa', 'disuruh', 'pinjol', 'biadab', 'bangun', 'ikn', 'korupsi', 'bts', 'biayain', 'food', 'estate', 'biayain', 'anak', 'bangsa', 'kagak']
['jika', 'dia', 'terpilih', 'gue', 'rasa', 'projek', 'ikn', 'juga', 'kemungkinan', 'akan', 'dibatalkan', 'layaknya', 'pulau', 'reklamasi', 'kemungkinan', 'karena', 'ikn', 'projek', 'besar', 'banget', 'dan', 'ambisius', 'pasti', 'akan', 'menantang', 'banget', 'untuk', 'membataalkannya']	['terpilih', 'projek', 'ikn', 'dibatalkan', 'layaknya', 'pulau', 'reklamasi', 'ikn', 'projek', 'banget', 'ambisius', 'menantang', 'banget', 'membataalkannya']

3.3.3.5 Stemming

Stemming mengubah kata ke bentuk dasar untuk menghilangkan infleksi, seperti akhiran dan imbuhan, sehingga berbagai bentuk kata diwakili oleh satu kata dasar. Stemming menyederhanakan teks dan meningkatkan efisiensi serta akurasi dalam pemrosesan bahasa alami, seperti pencarian dan klasifikasi teks. Diagram alir Stemming ditunjukkan pada Gambar 3.9.



Gambar 3.9 Alur proses *stemming*

Sebagai ilustrasi dari proses stemming, contoh *tweet* yang dihasilkan ditampilkan pada tabel 3.6.

Tabel 3.6 Contoh hasil proses *stemming*

<i>Stemming</i>	
Sebelum	Sesudah
['tugas', 'pemerintah', 'utama', '2', 'konstitusi', 'mencerdaskan', 'kebidupan', 'rakyat', 'mensejahterakan', 'rakyat', 'mahasiswa', 'disuruh', 'pinjol', 'biadab', 'bangun', 'ikn', 'korupsi', 'bts', 'biayain', 'food', 'estate', 'biayain', 'anak', 'bangsa', 'kagak']	tugas perintah utama 2 konstitusi cerdas kebidupan rakyat sejahtera rakyat mahasiswa suruh pinjol biadab bangun ikn korupsi bts biayain food estate biayain anak bangsa kagak

['terpilih', 'projek', 'ikn', 'dibatalkan', 'layaknya', 'pulau', 'reklamasi', 'ikn', 'projek', 'banget', 'ambisius', 'menantang', 'banget', 'membatalkannya']	pilih projek ikn batal layak pulau reklamasi ikn proyek banget ambisius tantang banget batal
---	--

3.3.1 Labeling

Labeling menggunakan TextBlob adalah proses memberikan label pada teks berdasarkan analisis sentimen. TextBlob adalah library Python yang memanfaatkan teknik pemrosesan bahasa alami (NLP) untuk menganalisis teks melalui dua metrik utama: polarity (nilai sentimen dari -1 untuk sangat negatif hingga 1 untuk sangat positif) dan subjectivity (tingkat subjektivitas dari 0 untuk sangat objektif hingga 1 untuk sangat subjektif). TextBlob sangat mudah digunakan, cepat, dan cocok untuk pemula, terutama untuk teks berbahasa Inggris.

3.3.2 Estraksi Fitur

Tahap ekstraksi fitur menggunakan algoritma TF-IDF mengubah teks menjadi vektor dengan memberi bobot pada setiap kata berdasarkan frekuensi dan kepentingannya dalam koleksi dokumen. Pembobotan ini menghasilkan representasi teks yang lebih informatif, mendukung proses klasifikasi atau analisis lanjutan, dan digunakan sebagai input pada proses klasifikasi.

Tabel 3.7 Contoh Dokumen *Tweet* Data Uji

D1	jokowi usaha singapura minat investasi ikn
D2	taksi terbang ikn buat hyundai uji coba juli
D3	percaya ikn nggak mangkrak

Proses pembobotan kata dimulai dengan mengambil kata-kata yang telah melalui tahap *Preprocessing* dari database. Kemudian, frekuensi kemunculan setiap kata dalam tiap *tweet* dihitung. Perhitungan ini menggunakan kata-kata yang telah melalui tahap *Preprocessing* dan menghasilkan bobot untuk setiap kata, seperti yang ditunjukkan pada tabel 3.8 berikut.

Tabel 3.8 Hasil perhitungan TF-IDF

TERM	TF			DF	idf = $\log \left(\frac{N}{DF} \right)$	TF*IDF		
	D1	D2	D3			D1	D2	D3
jokowi	1	0	0	1	0.477	0.477	0	0
usaha	1	0	0	1	0.477	0.477	0	0
singapura	1	0	0	1	0.477	0.477	0	0
minat	1	0	0	1	0.477	0.477	0	0
investasi	1	0	0	1	0.477	0.477	0	0
ikn	1	1	1	3	0	0	0	0
taksi	0	1	0	1	0.477	0	0.477	0
terbang	0	1	0	1	0.477	0	0.477	0
buat	0	1	0	1	0.477	0	0.477	0
hyundai	0	1	0	1	0.477	0	0.477	0
uji	0	1	0	1	0.477	0	0.477	0
coba	0	1	0	1	0.477	0	0.477	0
juli	0	1	0	1	0.477	0	0.477	0
percaya	0	0	1	1	0.477	0	0	0.477
nggak	0	0	1	1	0.477	0	0	0.477
mangkrak	0	0	1	1	0.477	0	0	0.477

3.3.3 Klasifikasi *Naïve Bayes*

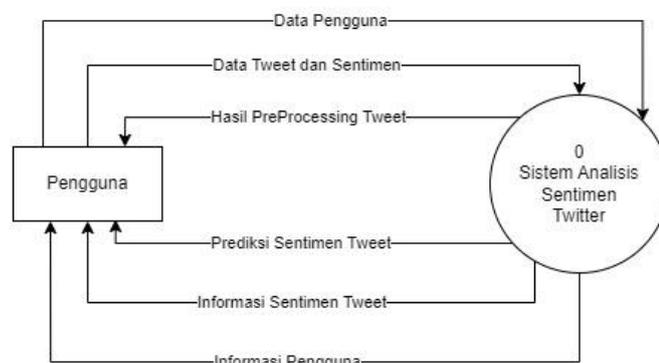
Setelah data ulasan diubah menjadi vektor, dataset kemudian dibagi menjadi dua bagian menggunakan fungsi `train_test_split` dari library Python. Bagian pertama, yang merupakan data latih, digunakan untuk melatih model. Setelah model dilatih menggunakan algoritma *Naïve Bayes*, langkah selanjutnya adalah menguji model tersebut dengan data uji untuk melakukan klasifikasi sentimen. Algoritma *Naïve Bayes* digunakan untuk menghitung nilai teks berdasarkan topik dan sentimen. Kata-kata yang telah melalui tahap Preprocessing dan dihitung frekuensi kemunculannya, seperti yang terlihat pada tabel 3.8, digunakan sebagai input awal. Setelah fitur kemunculan kata diperoleh, bobot kata terhadap topik dan sentimen dihitung. Bobot yang digunakan untuk menentukan topik dan sentimen adalah nilai tertinggi dari berbagai kemungkinan yang telah dihitung. Dengan menghitung probabilitas tweet menggunakan metode *Naïve Bayes*, dapat diketahui bahwa teks X T1 termasuk kategori positif karena bobot probabilitasnya lebih tinggi dibandingkan kategori netral dan negatif. Demikian pula, teks X T2 termasuk

kategori negatif karena bobot probabilitasnya lebih besar dibandingkan kategori lainnya, sedangkan teks X T3 termasuk kategori netral karena bobot probabilitasnya untuk kategori netral lebih tinggi dibandingkan kategori lainnya. Proses pengujian ini menghitung probabilitas dan memilih probabilitas tertinggi. Hasil perhitungan data uji ditunjukkan pada tabel 3.9.

Tabel 3.9 Contoh Hasil data uji

	<i>Tweet</i>	Probability Netral	Probability Positif	Probability Negatif	Predicted Label
T1	ikn menjadi pusat bisnis baru	0.284607	0.405617	0.309776	Positif
T2	tidak ada yang mau investasi di ikn	0.229561	0.283921	0.486518	Negatif
T3	pengembangan ikn sedang dalam proses	0.453955	0.260922	0.285123	Netral

3.4 Diagram Konteks



Gambar 3.10 Diagram Konteks

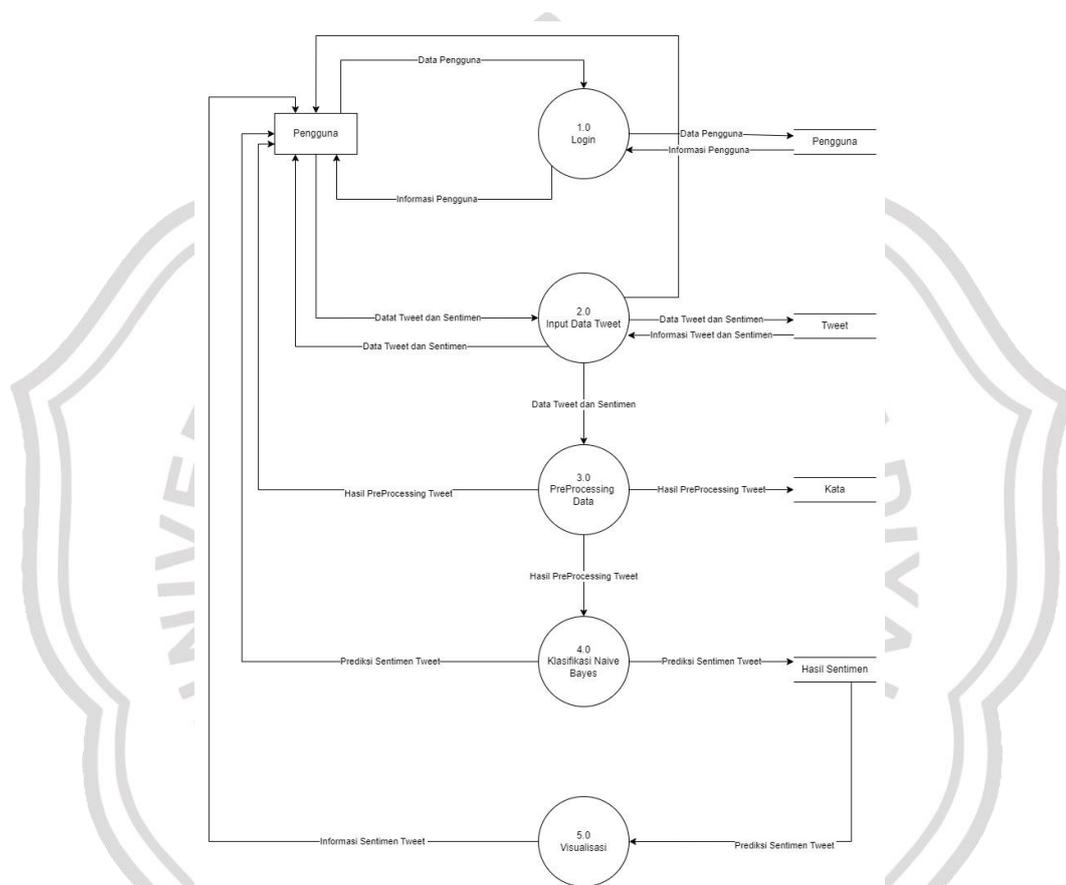
Diagram konteks ini menunjukkan alur interaksi antara pengguna dan sistem analisis sentimen X. Pengguna memasukkan data profil dan *tweet* ke dalam sistem, yang kemudian diproses untuk menghasilkan prediksi sentimen. Hasil *Preprocessing* dan prediksi sentimen ini dikembalikan kepada pengguna, bersama dengan informasi tambahan terkait sentimen *tweet* dan data pengguna. Diagram ini

merangkum bagaimana pengguna berinteraksi dengan sistem dari input hingga penerimaan hasil analisis.

3.5 Data Flow Diagram

3.5.1 Data Flow Diagram level 1

a) DFD Level 1 Proses 2



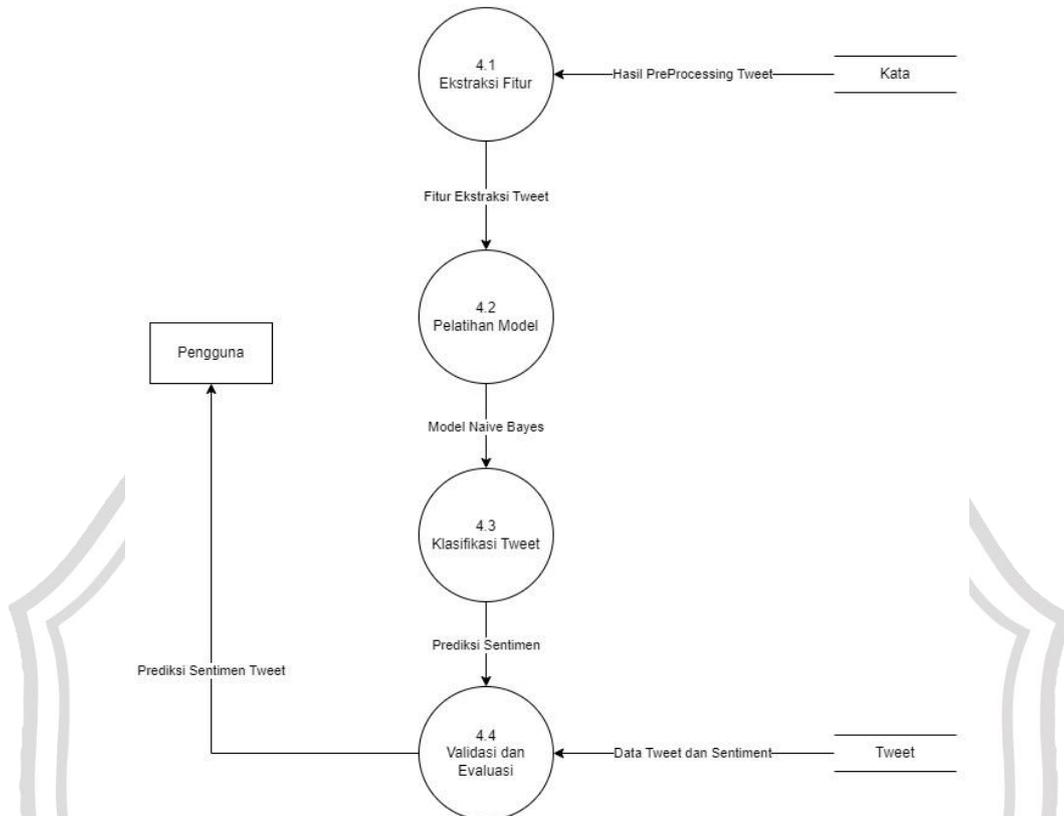
Gambar 3.11 Data Flow Diagram level 1 proses 2

Berdasarkan pada gambar 3.11 dapat dijelaskan bahwa DFD Level 1 proses 2 menjelaskan tiga proses detail dari proses klasifikasi yakni:

1. Proses 1.0 Login:
 - a. Pengguna: User atau pengguna aplikasi.
 - b. Pengguna memasukkan data login (misalnya username dan password) ke dalam sistem.

- c. Sistem memverifikasi data pengguna dan memberikan informasi yang diperlukan kembali ke pengguna (seperti notifikasi berhasil login atau gagal).
2. Proses 2.0 Input Data *Tweet*:
 - a. Setelah login, pengguna dapat memasukkan data *tweet* dan sentimen ke dalam sistem.
 - b. Data *tweet* dan sentimen ini dapat diinput secara manual oleh pengguna atau diambil dari sumber tertentu.
 - c. Proses ini menyimpan data *tweet* dan sentimen untuk diproses lebih lanjut dalam sistem.
 3. Proses 3.0 *Preprocessing* Data:
 - a. Data *tweet* yang diinput akan diproses dalam tahap *Preprocessing*.
 - b. Proses ini mencakup langkah-langkah seperti tokenisasi, penghapusan stopwords, stemming, dan lain-lain untuk mengubah *tweet* menjadi kata-kata yang siap dianalisis.
 - c. Hasil dari *Preprocessing* ini adalah data *tweet* yang sudah bersih dan siap digunakan untuk analisis lebih lanjut.
 4. Proses 4.0 Klasifikasi *Naive Bayes*:
 - a. Data hasil *Preprocessing* kemudian dimasukkan ke dalam algoritma *Naive Bayes* untuk diklasifikasikan.
 - b. Algoritma ini akan memprediksi sentimen *tweet* berdasarkan data yang telah diinput, apakah positif, negatif, atau netral.
 - c. Hasil prediksi sentimen kemudian diteruskan untuk divisualisasikan.
 5. Proses 5.0 - Visualisasi:
 - a. Prediksi sentimen yang dihasilkan dari proses klasifikasi ditampilkan dalam bentuk visualisasi kepada pengguna.
 - b. Visualisasi ini bisa berupa grafik, diagram, atau bentuk representasi data lainnya yang membantu pengguna memahami distribusi sentimen dari *tweet* yang dianalisis.

b) DFD Level 1 Proses 3



Gambar 3.12 Data Flow Diagram level 1 proses 3

Berdasarkan pada gambar 3.12 dapat dijelaskan bahwa DFD Level 1 proses 3 menjelaskan tiga proses detail dari proses klasifikasi yakni:

1. Proses 4.1 - Ekstraksi Fitur:
 - a. Input: Hasil *Preprocessing Tweet*, yaitu data *tweet* yang sudah dibersihkan dan dipersiapkan untuk analisis.
 - b. Output: Proses ini mengekstraksi fitur-fitur yang relevan dari *tweet*, seperti kata-kata penting, frekuensi kata, dan lain sebagainya, yang kemudian akan digunakan dalam pelatihan model.
 - c. Fitur-fitur ini merupakan representasi dari *tweet* dalam bentuk yang dapat diproses oleh algoritma *Naive Bayes*.

2. Proses 4.2 - Pelatihan Model:

- a. Input: Fitur Ekstraksi *Tweet*, yaitu data yang telah diekstraksi pada proses sebelumnya.
- b. Output: Model *Naive Bayes* yang terlatih. Pada proses ini, algoritma *Naive Bayes* dilatih menggunakan data fitur untuk memahami pola-pola yang ada dalam data dan memprediksi sentimen *tweet*.
- c. Model yang dihasilkan akan digunakan pada proses klasifikasi.

3. Proses 4.3 - Klasifikasi *Tweet*:

- a. Input: Model *Naive Bayes* yang telah dilatih dan fitur dari *tweet* yang baru.
- b. Output: Prediksi sentimen dari *tweet*, yang bisa berupa sentimen positif, negatif, atau netral.
- c. Proses ini menggunakan model yang telah dilatih untuk mengklasifikasikan *tweet* baru dan memberikan prediksi sentimennya.

4. Proses 4.4 - Validasi dan Evaluasi:

- a. Input: Prediksi Sentimen dan Data *Tweet* dan Sentimen yang sudah ada.
- b. Output: Hasil validasi dan evaluasi yang bertujuan untuk mengukur seberapa akurat prediksi sentimen yang dihasilkan oleh model.
- c. Proses ini membandingkan hasil prediksi dengan data sentimen yang sudah diketahui untuk mengevaluasi kinerja model.

3.6 Entity Relationship Diagram (ERD)

Entity Relationship Diagram (ERD) adalah model konseptual yang menggambarkan keterkaitan antar tabel dalam sistem, memodelkan struktur data dan relasi antar data. Dalam sistem analisis sentimen Ibu Kota Nusantara, ERD merancang struktur data dan relasi antar entitas. Desain ERD ini dapat dilihat pada Gambar 3.13.

users		training_datasets		test_datasets	
id	integer	id	integer	id	integer
name	varchar	full_text	text	full_text	text
username	varchar	label	varchar	expected_sentiment	varchar
password	varchar	preprocessed_text	text	predicted_sentiment	varchar
remember_token	varchar	created_at	timestamp	created_at	timestamp
created_at	timestamp	updated_at	timestamp	updated_at	timestamp
updated_at	timestamp				

Gambar 3.13 Entity relationship diagram dalam analisis sentimen IKN

3.7 Perancangan Antarmuka

a) Antar Muka Halaman *Login*

Halaman login merupakan halaman pertama dari sistem, di mana admin perlu memasukkan username dan password yang telah terdaftar untuk dapat mengakses sistem. Tampilan desain halaman login dapat dilihat pada Gambar 3.14.

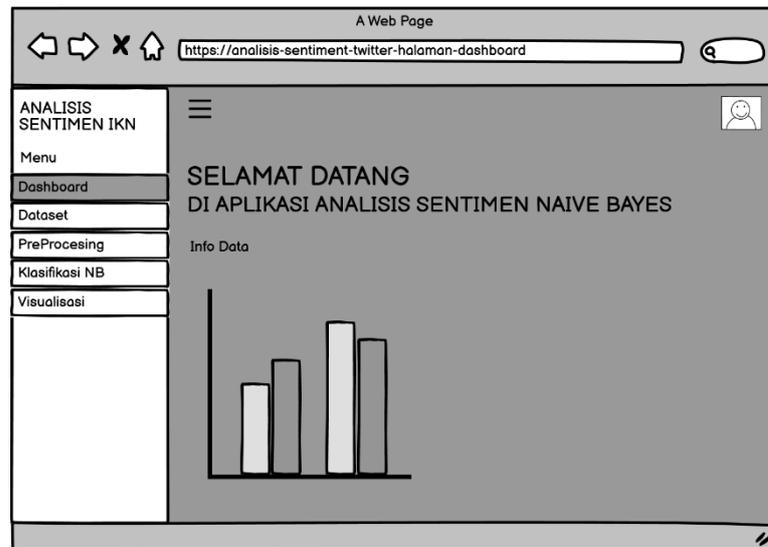
The image shows a web browser window with the following elements:

- Address bar: <https://analisis-sentiment-twitter>
- Page Title: A Web Page
- Form Title: ANALISIS SENTIMEN IBU KOTA NUSANTARA
- Form Fields:
 - USERNAME:
 - PASSWORD:
- Form Button: LOGIN

Gambar 3.14 Tampilan Halaman *Login*

b) Antar Muka Halaman Utama

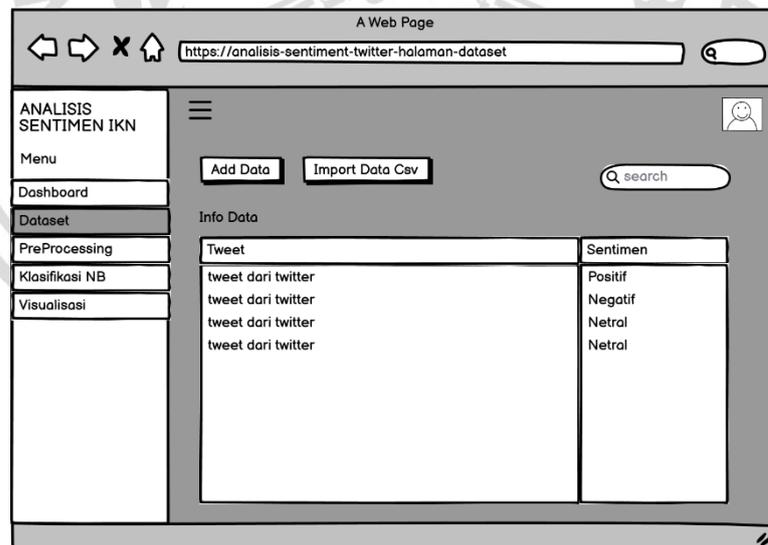
Halaman utama adalah halaman pertama yang muncul setelah pengguna berhasil *login*. Halaman ini menampilkan informasi mengenai nama sistem dan fungsinya. Tampilan rancangan halaman home dapat dilihat pada Gambar 3.15.



Gambar 3.15 Tampilan Halaman Utama

c) Antar Muka Halaman Dataset

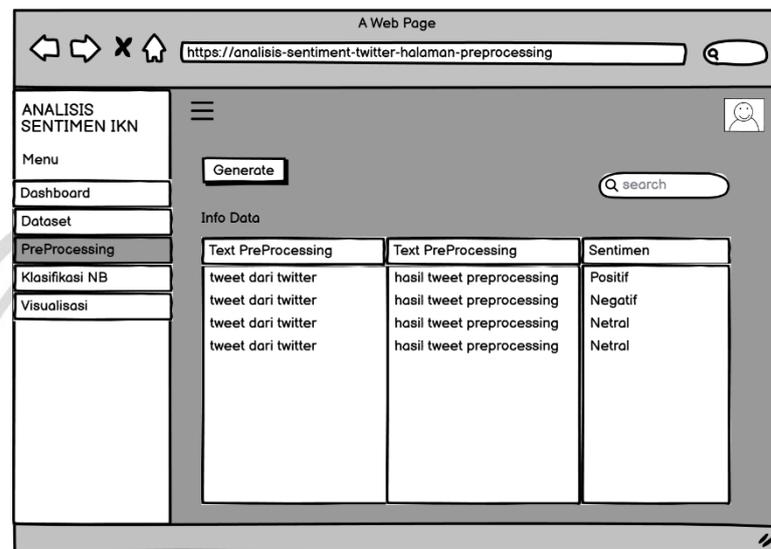
Halaman Dataset menampilkan data untuk analisis sentimen, di mana admin dapat melihat, menambah, mengubah, atau menghapus data tweet. Data mencakup teks tweet dan label sentimen. Fitur pencarian dan filter disediakan untuk memudahkan pengelolaan data. Tampilan halaman ini ada pada Gambar 3.16.



Gambar 3.16 Tampilan Halaman Dataset

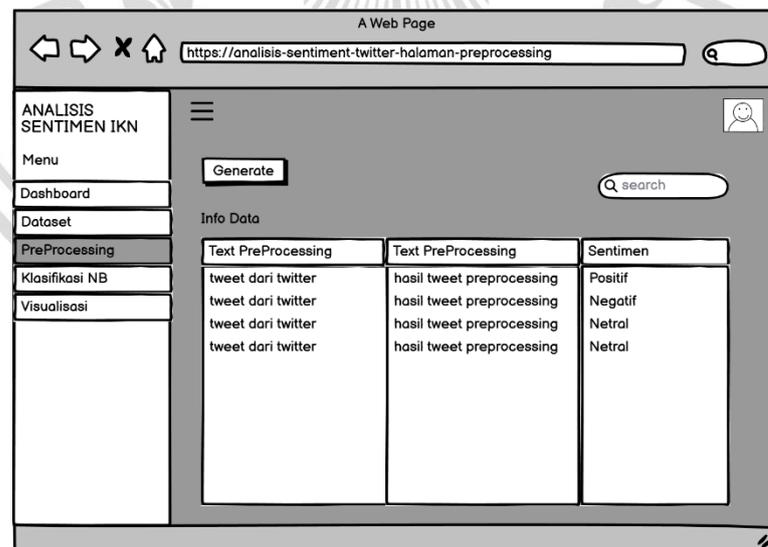
d) Antar Muka Halaman *Preprocessing*

Halaman Preprocessing digunakan untuk memproses data mentah tweet sebelum analisis. Tujuannya adalah membersihkan dan menyiapkan data agar siap untuk analisis sentimen menggunakan metode Naive Bayes Classifier, yang penting untuk meningkatkan akurasi. Tampilan halaman ini ada pada Gambar 3.17.



Gambar 3.17 Tampilan Halaman *Preprocessing*

e) Antar Muka Halaman Klasifikasi

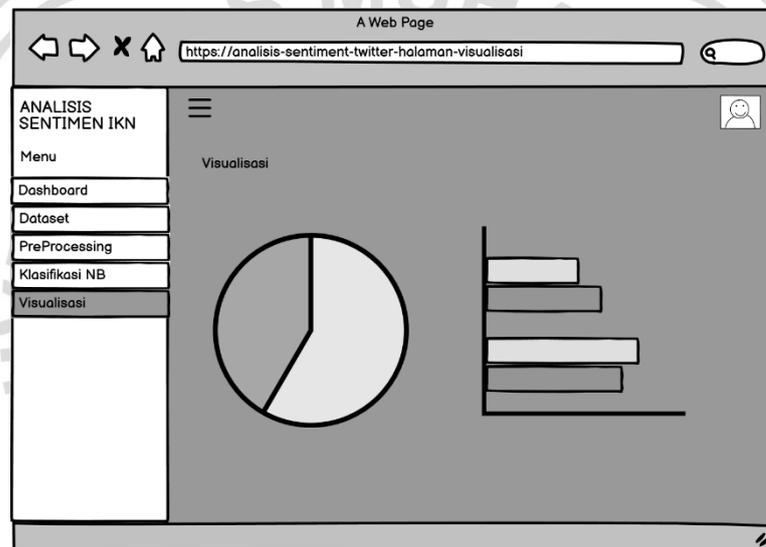


Gambar 3.18 Tampilan Halaman Klasifikasi

Halaman Klasifikasi mengklasifikasikan data tweet berdasarkan sentimen menggunakan metode Naive Bayes Classifier. Pengguna dapat memilih data yang telah dipreproses dan menjalankan klasifikasi untuk menentukan sentimen positif, negatif, atau netral, lengkap dengan persentase keakuratannya. Tampilan halaman ini terdapat pada Gambar 3.18.

f) Antar Muka Halaman Visualisasi

Halaman Visualisasi menampilkan hasil analisis sentimen dalam bentuk grafis untuk memudahkan pemahaman tren sentimen publik terhadap Ibu Kota Nusantara di media sosial X. Tampilan halaman ini terdapat pada Gambar 3.19.



Gambar 3.19 Tampilan Halaman Visualisasi

3.8 Skenario Pengujian

Pengujian dalam pembuatan sistem analisis perlu dilakukan untuk mengetahui tingkat akurasi perhitungan dengan metode Naïve Bayes Classifier. Dalam penelitian ini, pengujian dilakukan dengan menggunakan Confusion Matrix. Skenario pengujian ini bertujuan untuk menentukan analisis sentimen dan menghitung nilai akurasi dari sistem yang telah dibuat. Penulis akan membagi dataset yang berjumlah 4180 *tweet* dengan rasio 70:30% untuk melihat akurasi yang diberikan oleh model. Adapun rasio pembagian tersebut disajikan pada tabel 3.10.

Tabel 3.10 Pembagian dataset

Rasio Pembagian % Data Latih : Data Uji	Jumlah Data Latih	Jumlah Data Uji
70 : 30	2926	1254

3.9 Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja model klasifikasi. Proses ini mencakup beberapa langkah, termasuk penggunaan confusion matrix, tabel akurasi, serta perhitungan presisi untuk setiap model yang dievaluasi. Setelah menguji data uji dengan model yang telah dilatih menggunakan data latih, prediksi kelas untuk setiap sampel dalam data uji akan dihasilkan. Prediksi kelas tersebut kemudian dibandingkan dengan kelas yang sebenarnya, yang sudah ditandai sebelumnya pada data uji yang tersembunyi. Evaluasi ini menghasilkan nilai akurasi, presisi, recall, dan F1-score, yang memberikan gambaran tentang seberapa baik model dalam mengklasifikasikan data uji. Untuk menilai kinerja Algoritma Naïve Bayes, pengujian terhadap model dilakukan, dan hasil klasifikasi ditampilkan dalam bentuk confusion matrix. Tabel confusion matrix ini mencakup kelas prediksi dan kelas aktual. Model confusion matrix dengan ukuran 3x3 dapat dilihat pada Tabel 3.11.

Tabel 3.11 Tabel hasil *confusion matrix* pada data uji

<i>True Class</i>	<i>Predicted Class</i>		
	Negatif	Netral	Positif
Negatif	221	77	256
Netral	1	836	215
Positif	0	22	1298

Tabel 3.12 Tabel hasil keseluruhan evaluasi model

	(Rasio 70:30)
<i>Accuracy (%)</i>	80%
<i>Precision (%)</i>	87%
<i>Recall (%)</i>	73%
<i>F1-Score (%)</i>	75%

Tabel Evaluasi pada gambar menunjukkan hasil evaluasi kinerja model klasifikasi dengan rasio data *training* dan *testing* 70:30. Berdasarkan tabel tersebut, model mencapai akurasi sebesar 80%, yang menunjukkan bahwa 80% dari seluruh prediksi yang dibuat oleh model adalah benar. *Precision* dari model adalah 87%, yang berarti dari seluruh data yang diprediksi positif oleh model, 87% di antaranya benar-benar positif. *Recall* sebesar 73% menunjukkan bahwa dari seluruh data yang sebenarnya positif, model berhasil memprediksi dengan benar sebanyak 73%. Sementara itu, *F1-Score* yang diperoleh adalah 75%, mencerminkan keseimbangan yang baik antara presisi dan recall, sehingga model ini dapat dianggap cukup handal dalam memprediksi baik kategori positif maupun negatif