

Analisis Sentimen Opini Publik Terhadap Kunjungan Paus Fransiskus di Indonesia Menggunakan Algoritma *Naïve Bayes*

Nadia S. Sugi¹, Lanny Sitanayah^{1*}, Junaidy B. Sanger¹

¹*Program Studi Teknik Informatika, Fakultas Teknik, Universitas Katolik De La Salle Manado, Manado, Indonesia
Email:lsitanayah@unikadelasalle.ac.id**

Abstract— The X social media platform (formerly known as Twitter) has become one of the main media for discussing various issues and public opinions, including Pope Francis' visit to Indonesia on September 3-6, 2024. This study aims to analyze public opinion sentiment based on tweet data by applying the Naïve Bayes algorithm. Data collection was carried out through crawling techniques using the Twitter API, which was run on Google Colab. The Naïve Bayes algorithm is used to classify data into three types of sentiment: positive, negative, and neutral. Testing was conducted using training and testing data, and then evaluated through a confusion matrix to measure accuracy, precision, recall, and F1-score. The analysis results show that public opinion regarding the visit is positive, with an evaluation yielding average accuracy metric results, namely the 90:10 ratio reaching 74.37%, followed by the 80:20 ratio with 73.26%, the 70:30 ratio with 70.91%, and the 60:40 ratio with 66.44%. The best precision, recall, and F1-score metrics were obtained from the 90:10 ratio, which were 59.72%, 59.29%, and 59.48%, respectively. This research provides an overview of public perception regarding Pope Francis' visit and demonstrates the results of applying the Naïve Bayes algorithm. Thus, this application is expected to contribute to the advancement of sentiment analysis methods in public opinion analysis on various social issues.

Keywords— Sentiment Analysis, Public Opinion, Naïve Bayes, Pope Francis' Visit, X Social Media Platform

Abstrak— Platform media sosial X (sebelumnya dikenal sebagai Twitter) menjadi salah satu media utama untuk membahas berbagai isu dan opini publik, termasuk kunjungan Paus Fransiskus di Indonesia pada tanggal 3-6 September 2024. Penelitian ini bertujuan untuk menganalisis sentimen opini publik berdasarkan data *tweet* dengan menerapkan algoritma *Naïve Bayes*. Pengumpulan data dilakukan melalui teknik *crawling* menggunakan Twitter API yang dijalankan pada *Google Colab*. Algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan data ke dalam tiga jenis sentimen, yaitu positif, negatif, dan netral. Pengujian dilakukan dengan memanfaatkan data latih dan data uji, kemudian dievaluasi melalui *confusion matrix* untuk mengukur akurasi, presisi, *recall*, dan *F1-score*. Hasil analisis menunjukkan bahwa opini publik terhadap kunjungan tersebut bersifat positif, dengan evaluasi diperoleh hasil rata-rata metrik akurasi, yaitu rasio 90:10 mencapai 74,37%, kemudian rasio 80:20 dengan 73,26%, rasio 70:30 dengan 70,91%, dan rasio 60:40 dengan 66,44%. Metrik presisi, *recall*, dan *F1-score* terbaik diperoleh dari rasio 90:10 yaitu 59,72%, 59,29%, dan 59,48%. Penelitian ini memberikan gambaran mengenai persepsi publik terhadap kunjungan Paus Fransiskus dan memperlihatkan hasil penerapan algoritma *Naïve Bayes*. Dengan demikian, aplikasi ini diharapkan dapat berkontribusi dalam kemajuan metode analisis sentimen pada analisis opini publik di berbagai isu sosial.

Kata Kunci— Analisis Sentimen, Opini Publik, *Naïve Bayes*, Kunjungan Paus Fransiskus, Platform Media Sosial X

I. PENDAHULUAN

Inovasi dalam teknologi terus melaju seiring dengan perkembangan zaman yang memberikan perubahan pada berbagai bidang. Salah satu transformasi yang terjadi yaitu pada bidang hiburan seperti media sosial. Sebagai ruang digital, media sosial memungkinkan pengguna untuk dapat membina komunikasi, bertukar informasi dan menjalin komunikasi dengan orang lain dengan Internet [1]. Salah satu media sosial yang saat ini sangat populer adalah Twitter yang sudah berubah nama menjadi X.

X merupakan suatu aplikasi media sosial yang dulunya memiliki nama Twitter, namun sudah resmi berubah nama sejak Juli 2023 setelah diakuisisi perusahaannya oleh Elon Musk. X saat ini dikenal masyarakat karena memiliki kemampuan *microblogging* yang bisa membuat pengguna

membagikan informasi, pemikiran, tautan, gambar dan video yang disebut dengan “*Tweet*” [2]. X dapat membantu pengguna dalam memberikan dan menyampaikan informasi, pemikiran, reaksi, emosi, dan sudut pandang dari pengguna kepada publik terhadap isu-isu terkini yang menjadi perhatian publik.

Salah satu isu yang menjadi perhatian publik, yaitu mengenai kedatangan Paus Fransiskus di Indonesia pada September 2024. Paus Fransiskus merupakan pemimpin Vatikan dan otoritas tertinggi gereja Katolik sedunia dan sering melakukan perjalanan apostolik pada beberapa negara, salah satunya di Indonesia pada tanggal 3-6 September 2024 yang lalu [3]. Kedatangan Paus menjadi topik yang sering dibahas dalam berbagai media sosial, salah satunya pada media sosial X. Kunjungan Paus ke Indonesia ini tentunya memiliki perspektif yang berbeda-beda pada pandangan setiap orang. Ada yang menanggapi

dengan positif, netral, namun ada juga yang memiliki tanggapan negatif. Oleh karena itu, perlu adanya penelitian analisis sentimen yang berguna untuk dapat memberikan informasi terkait tanggapan, emosi, reaksi dan pandangan masyarakat mengenai kunjungan Paus Fransiskus ke Indonesia sehingga dapat mengetahui persepsi publik dan dampak sosial bagi masyarakat Indonesia.

Analisis sentimen adalah metode untuk mengelompokkan dokumen teks menjadi tiga kategori positif, netral, dan negatif. Analisis sentimen menjadi salah satu sub-bidang dari *data mining* yang dipakai dengan tujuan menganalisis serta menyaring informasi dari teks yang mencakup pendapat, penilaian, dan emosi seseorang terhadap produk atau isu tertentu [4]. Terdapat berbagai jenis algoritma yang digunakan pada analisis sentimen, salah satunya yaitu algoritma *Naïve Bayes*. *Naïve Bayes* dikenal sebagai sebuah algoritma atau metode klasifikasi probabilitas yang ringkas, yang menentukan kemungkinan dengan menjumlahkan frekuensi termasuk perpaduan nilai dari data yang tersedia [5].

Naïve Bayes sering digunakan dalam analisis teks dan pemrosesan bahasa alami [6]. Berdasarkan Teorema Bayes, algoritma ini mengasumsikan bahwa setiap fitur dalam data bersifat independen, sehingga cepat dan mudah diimplementasikan. Dalam analisis sentimen, *Naïve Bayes* dapat mengklasifikasikan teks ke dalam kategori seperti positif, negatif, atau netral. Meskipun sederhana, *Naïve Bayes* telah terbukti efektif dalam berbagai aplikasi analisis teks, termasuk ulasan produk, komentar media sosial, dan umpan balik pengguna, sehingga menjadikannya pilihan populer untuk penelitian yang melibatkan data teks yang besar dan kompleks.

Terdapat beberapa riset seputar analisis sentimen menggunakan algoritma *Naïve Bayes*. Misalnya, sebuah penelitian yang menganalisis sentimen terhadap kandidat presiden Indonesia 2019 di Twitter mencapai akurasi 80,90% menggunakan *Naïve Bayes* dengan data pelatihan untuk setiap kandidat [7]. Penelitian ini juga membandingkan *Naïve Bayes* dengan *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (K-NN), menemukan bahwa *Naïve Bayes* mencapai akurasi tertinggi sebesar 75,58%. Kemudian, penelitian lain yaitu perbandingan algoritma *Naïve Bayes* dan SVM pada analisis sentimen *Spotify*. Pada penelitian ini, dilakukan uji coba dan perbandingan kedua algoritma ini, di mana nilai akurasi dijadikan sebagai patokan penilaian pada kasus analisis sentimen ini. Hasil menunjukkan tingkat akurasi tertinggi diperoleh pada algoritma *Naïve Bayes*, yaitu 86,4% dibandingkan dengan algoritma SVM, yaitu 84% [8]. Adapun penelitian lainnya, yaitu perbandingan metode K-NN, *Naïve Bayes*, dan *Decision Tree* untuk analisis sentimen *tweet* Twitter terkait opini terhadap PT PAL Indonesia. Hasil menunjukkan *Naïve Bayes* memiliki tingkat akurasi tertinggi dibandingkan kedua algoritma lainnya sebesar 84,08% [9]. Terdapat juga penelitian lainnya, yaitu analisis sentimen pengguna aplikasi *CapCut* pada ulasan di *Playstore* menggunakan metode *Naïve Bayes*. Hasil akurasi metode *Naïve Bayes* mencapai nilai 84,09% [10]. Hasil dari penelitian tersebut menyimpulkan

bahwa Algoritma *Naïve Bayes* efektif diterapkan dalam kasus analisis sentimen.

Dengan merujuk pada uraian masalah yang telah dipaparkan sebelumnya, maka dibutuhkan sebuah aplikasi analisis sentimen agar pengguna bisa mendapatkan informasi terkait pandangan masyarakat terhadap kunjungan Paus Fransiskus di Indonesia. Oleh karena itu, akan dibangun suatu aplikasi *online* dengan menerapkan algoritma *Naïve Bayes* dalam menganalisis sentimen opini publik terhadap kunjungan Paus Fransiskus di Indonesia pada media sosial X untuk dapat memberikan informasi terkait tanggapan, emosi, reaksi dan pandangan masyarakat mengenai kunjungan Paus Fransiskus di Indonesia, sehingga dapat mengetahui persepsi publik dan dampak sosial bagi masyarakat Indonesia.

II. TINJAUAN PUSTAKA

Analisis sentimen merupakan pendekatan guna mengidentifikasi dan menggolongkan emosi, atau opini yang terdapat pada teks, misalnya ulasan, postingan, atau artikel berita. Teknik ini bertujuan untuk menilai apakah teks yang dianalisis memiliki sentimen positif, negatif, atau netral, serta dapat diterapkan pada berbagai data teks untuk memperoleh wawasan dari perspektif pengguna atau publik [11].

Analisis sentimen adalah alat yang krusial bagi berbagai sektor untuk memahami pandangan publik, mendapatkan wawasan tentang persepsi pelanggan, dan mendukung pengambilan keputusan yang didasarkan pada data. Dengan kemajuan dalam teknik *Natural Language Processing* (NLP), analisis sentimen saat ini mampu menghasilkan hasil yang lebih tepat dan berguna untuk kepentingan bisnis dan sosial [12].

Naïve Bayes adalah teknik prediksi yang memanfaatkan probabilitas sederhana menggunakan teorema *Bayes* dengan anggapan independensi yang kuat, di mana model fitur dianggap saling independen, asumsi independensi yang kuat berarti data tidak memiliki keterkaitan dalam konteks yang sama maupun atribut lainnya [13].

Dalam *Naïve Bayes*, masalah dikategorikan ke dalam kelas-kelas berdasarkan perbedaan dan persamaan karakteristik, dan statistik digunakan untuk memprediksi kemungkinan sebuah kelas terjadi. Algoritma ini sangat diminati karena mudah digunakan, efisien dalam memproses *dataset* besar, dan dapat beroperasi dengan baik meskipun terdapat asumsi independensi yang kuat antara fitur-fitur yang ada. Berikut ini adalah rumus *Naïve Bayes* [14].

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \dots\dots\dots (1)$$

Keterangan:

X : Data yang belum ditentukan kelasnya.

H : Data X dipandang sebagai kelas khusus dalam hipotesis ini.

$P(H|X)$: Peluang terjadinya hipotesis H dalam kondisi X (*posterior probability*).

$P(H)$: Peluang hipotesis H (*class prior probability*).

$P(X|H)$: Peluang X yang dihitung berdasarkan hipotesis H (*likelihood*).

$P(X)$: Peluang X dari keseluruhan data.

CountVectorizer adalah teknik pemrosesan teks yang dipakai untuk mengukur frekuensi kemunculan setiap kata pada suatu *file* serta mengubahnya dalam bentuk matriks (*term-document matrix*). Pada matriks ini, setiap baris merepresentasikan suatu dokumen dan setiap kolom menunjukkan kata unik dari keseluruhan kumpulan dokumen. Sebelum dianalisis oleh algoritma *machine learning*, *countvectorizer* biasanya sering digunakan sebagai tahap awal untuk mempersiapkan data teks, yang dimana teks mentah akan diubah menjadi data numerik sehingga dapat diproses lebih lanjut [15].

Dalam *machine learning*, evaluasi model atau algoritma dilakukan dengan bantuan *confusion matrix* yang membandingkan hasil prediksi sistem dan nilai aktual dalam suatu *dataset*. *Confusion matrix* terdiri dari empat bagian utama yang menggambarkan hubungan antara hasil prediksi dan nilai aktual. *Confusion matrix* dapat dilihat pada ilustrasi Gambar 1.

		PREDICTED	
		Positive	Negative
ACTUAL	Positive	TRUE POSITIVE	FALSE NEGATIVE
	Negative	FALSE POSITIVE	TRUE NEGATIVE

Gbr. 1 *Confusion Matrix* [16]

Keterangan:

1. TP: data aktual positif dan model prediksi menjadi data positif.
2. TN: data aktual negatif dan model prediksi menjadi data negatif.
3. FP: data aktual negatif, tapi model prediksi menjadi data positif.
4. FN: data aktual positif, tapi model prediksi menjadi data negatif.

Dari keempat data tersebut, *confusion matrix* berguna untuk mengevaluasi kinerja model yang telah dibangun. Berikut penjelasan dan rumusnya [16]:

1. *Accuracy*: tingkat keberhasilan model dalam membuat hasil prediksi yang benar dari seluruh data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots(2)$$

2. *Precision*: saat model membuat prediksi positif, seberapa akurat prediksi tersebut.

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots(3)$$

3. *Recall (sensitivity)*: saat data aktualnya positif, seberapa sering model memprediksi positif.

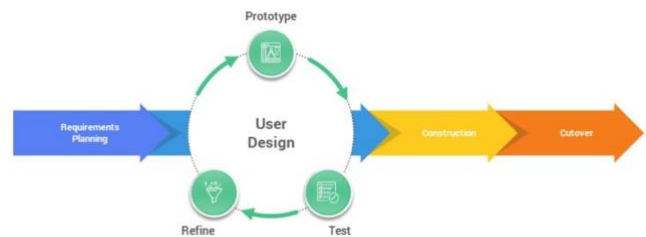
$$Recall = \frac{TP}{TP + FN} \dots\dots\dots(4)$$

4. *F1-Score*: keseimbangan antara *presicion* dan *recall (sensitivity)*.

$$F1-Score = 2 \times \frac{precision \times recall}{precision + recall} \dots\dots\dots(5)$$

III. METODOLOGI

Untuk membangun aplikasi analisis sentimen, *Rapid Application Development* (RAD) diterapkan sebagai metodologi dalam penelitian ini karena singkat dan berfokus pada kecepatan pembuatan dan pengembangan aplikasi [17]. *Rapid Application Development* (RAD) mempunyai 4 fase utama seperti yang dapat dilihat pada Gambar 2.



Gbr. 2 Tahapan *Rapid Application Development*

1. *Requirements Planning*

Pada fase ini, pengguna dan analis bekerja sama untuk mendiskusikan tujuan aplikasi atau sistem yang akan dibangun dan menetapkan persyaratan informasi yang relevan dengan tujuan tersebut.

2. *User Design*

Pada fase ini, informasi dikumpulkan dan didokumentasikan untuk keperluan analisis dan perancangan desain aplikasi.

3. *Construction Phase*

Fase ini membahas pembuatan rancangan sistem dengan lebih detail. Dalam fase ini juga dilaksanakan implementasi dari desain sistem yang telah dirancang.

4. *Cutover*

Pada fase ini, aplikasi yang dibangun diuji untuk memastikan fungsionalitasnya dan melakukan peralihan aplikasi pada pengguna.

IV. HASIL DAN PEMBAHASAN

Berikut dijelaskan hasil dan pembahasan dari penelitian ini.

A. *Perhitungan Algoritma*

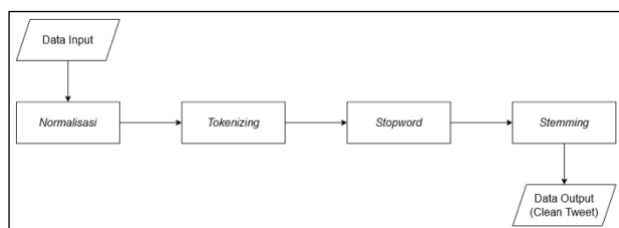
Tahapan ini menggambarkan proses algoritma dalam aplikasi yang terdiri dari beberapa tahap, yaitu:

Langkah pertama, pada aplikasi akan diunggah data hasil crawling dalam format CSV yang diperoleh melalui proses crawling data tweet. Informasi yang diambil yaitu data username dan data tweet. Berikut ini merupakan hasil crawling data dengan contoh sampel sebanyak 6 data.

TABEL I. CONTOH HASIL *CRAWLING*

No.	Username	Data Tweet
1.	Faizuldgn33	Kunjungan Paus Fransiskus ke Indonesia adalah wujud nyata dari semangat pluralitas di mana dua agama besar dunia bertemu untuk memperkuat komitmen terhadap perdamaian dan keadilan sosial. JagaKEBERAGAMANJagaKE UTUHAN #PausFransiskus #Pluralitas #IndonesiaDamai https://t.co/KNCxlelkz0
2.	Anneswaruwu	ingat bagimu agama mu bagiku agamaku. se simpel itu!!! beliau paus fransiskus datang juga dengan tujuan perdamaian. ga jelas lu asli ngomong kek gini.
3.	Kittiw123	Ngeliat fyp di tiktok tentang paus fransiskus jujur gatau itu siapa dan gmn tp kayak terharuu ngeliat betapaa baiknya bapa suci itu.
4.	Zestromx	agak iri sm orang yg ketemu langsung sama Paus Fransiskus.
5.	b_chocomint27	njir paus fransiskus katanya harus di deportasi ada aja manusia yaampun
6.	couwsun	Beli perangko paus fransiskus dmn ya

Langkah selanjutnya yaitu tahapan *pre-processing* seperti yang dapat dilihat pada Gambar 3. Data yang telah didapatkan dari proses *crawling* sebelumnya belum bisa diolah karena data tersebut masih memiliki karakter khusus, huruf duplikat, angka, dan emoji yang tidak diperlukan, oleh karena itu tahap *pre-processing* ini diperlukan untuk pembersihan data agar data lebih bersih dan terstruktur. Proses *pre-processing* dilakukan dalam beberapa teknik pemrosesan teks seperti *normalisasi*, *tokenisasi*, *stopword*, dan *stemming*.

Gbr. 3 Tahapan *Pre-processing*

1. *Normalisasi*
Proses ini mencakup *case folding* yang mengonversi menjadi huruf kecil, menghapus huruf berulang dan membersihkan simbol/karakter.
2. *Tokenisasi*
Setiap kata pada suatu kalimat dalam data *tweet* akan dipilah menjadi token.

3. *Stopword*
Menghapus kata yang sering muncul tetapi tidak relevan.
4. *Stemming*
Pada data *tweet* akan diubah kata yang berimbuhan menjadi kata dasar.

Selain *pre-processing*, akan dilakukan juga proses pelabelan manual data *tweet*. Berikut ini merupakan hasil *pre-processing* pada sampel data yang berjumlah 6 data.

TABEL II. CONTOH HASIL *PRE-PROCESSING*

No.	Data Tweet	Clean Tweet	Label
1.	Kunjungan Paus Fransiskus ke Indonesia adalah wujud nyata dari semangat pluralitas di mana dua agama besar dunia bertemu untuk memperkuat komitmen terhadap perdamaian dan keadilan sosial. JagaKEBERAGAMANJagaKE UTUHAN #PausFransiskus #Pluralitas #IndonesiaDamai https://t.co/KNCxlelkz0	kunjungan paus fransiskus wujud nyata semangat pluralitas besar dunia bertemu memperkuat komitmen terhadap perdamaian keadilan sosial	Positif
2.	ingat bagimu agama mu bagiku agamaku. se simpel itu!!! beliau paus fransiskus datang juga dengan tujuan perdamaian. ga jelas lu asli ngomong kek gini.	ingat bagimu agamamu bagiku agamaku paus fransiskus datang tujuan perdamaian ga asli ngomong gini	Negatif
3.	Ngeliat fyp di tiktok tentang paus fransiskus jujur gatau itu siapa dan gmn tp kayak terharuu ngeliat betapaa baiknya bapa suci itu.	ngeliat fyp tiktok paus fransiskus jujur gatau gimana terharu ngeliat baiknya suci	Positif
4.	agak iri sm orang yg ketemu langsung sama Paus Fransiskus.	iri sama ketemu langsung sama paus fransiskus	Netral
5.	njir paus fransiskus katanya harus di deportasi ada aja manusia yaampun	a****g paus fransiskus katanya harus deportasi ya	Negatif

No.	Data Tweet	Clean Tweet	Label
6.	Beli perangko paus fransiskus dmn ya	perangko paus fransiskus dimana ya	?

Langkah selanjutnya adalah pemisahan *dataset* dengan menggunakan teknik pembagian data. Proses ini dilakukan dengan menetapkan rasio pembagian data yang sesuai. Sebagai contoh, data 1 hingga 5 dijadikan data latih, sedangkan data 6 digunakan sebagai data uji.

TABEL III. CONTOH PEMISAHAN DATA

No.	Clean Tweet	Label	Jenis Data
1.	kunjungan paus fransiskus wujud nyata semangat pluralitas besar dunia bertemu memperkuat komitmen terhadap perdamaian keadilan sosial	Positif	Latih
2.	ingat bagimu agamamu bagiku agamaku paus fransiskus datang tujuan perdamaian ga asli ngomong gini	Negatif	Latih
3.	ngeliat fyp tiktok paus fransiskus jujur gatau Positif Latih gimana terharu ngeliat baiknya suci	Positif	Latih
4.	iri sama ketemu langsung sama paus fransiskus	Netral	Latih
5.	a****g paus fransiskus katanya harus deportasi ya	Negatif	Latih
6.	perangko paus fransiskus dimana ya	?	Uji

Langkah berikutnya adalah mengubah atau mengekstrak data *tweet* tersebut menjadi sebuah teks (fitur) sehingga format yang ada dapat dipahami oleh model pada aplikasi. Pada tahapan proses ekstraksi fitur ini, akan dilakukan metode *CountVectorizer* untuk menghitung frekuensi seberapa banyak munculnya sebuah kata dalam *dataset*.

Langkah terakhir adalah tahap *processing*, di mana analisis dimulai dengan penerapan algoritma. Prediksi sentimen dilakukan pada tingkat ulasan secara keseluruhan di aplikasi, meskipun perhitungan dilakukan pada kata-kata individu dalam ulasan tersebut. Setiap kata dianalisis untuk menentukan probabilitas kontribusinya terhadap sentimen. Probabilitas yang dihitung selanjutnya digabungkan agar memberikan prediksi sentimen akhir untuk ulasan yang baru.

Langkah awal dalam mengolah data latih dimulai dengan perhitungan probabilitas prior di setiap kelas. Berdasarkan Tabel 3, dari 5 data latih terdapat 2 *tweet* positif dan netral, serta 1 *tweet* negatif. Berikut ini merupakan perhitungan Probabilitas Prior:

$$P(\text{Positif}) = \frac{2}{5} = 0,4$$

$$P(\text{Negatif}) = \frac{2}{5} = 0,4$$

$$P(\text{Netral}) = \frac{1}{5} = 0,2$$

Selanjutnya, menghitung probabilitas *likelihood* dengan mengukur frekuensi kemunculan kata-kata di tiap kelas berdasarkan data latih, yang jumlahnya dipengaruhi oleh hasil *pre-processing*. Proses perhitungan probabilitas *likelihood* pada setiap kata di seluruh kelas dilakukan sebagai berikut:

Positif:

- $P(\text{kunjungan}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{paus}|\text{Positif}) = \frac{2}{25} = 0,08$
- $P(\text{fransiskus}|\text{Positif}) = \frac{2}{25} = 0,08$
- $P(\text{wujud}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{nyata}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{semangat}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{pluralitas}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{besar}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{dunia}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{bertemu}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{memperkuat}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{komitmen}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{terhadap}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{perdamaian}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{keadilan}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{sosial}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{ngeliat}|\text{Positif}) = \frac{2}{25} = 0,08$
- $P(\text{fyp}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{tiktok}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{jujur}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{gatau}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{gimana}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{terharu}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{baiknya}|\text{Positif}) = \frac{1}{25} = 0,04$
- $P(\text{suci}|\text{Positif}) = \frac{1}{25} = 0,04$

Negatif:

- $P(\text{ingat}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{bagimu}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{agamamu}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{bagiku}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{agamaku}|\text{Negatif}) = \frac{1}{19} = 0,053$

- $P(\text{paus}|\text{Negatif}) = \frac{2}{19} = 0,105$
- $P(\text{fransiskus}|\text{Negatif}) = \frac{2}{19} = 0,105$
- $P(\text{dateng}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{tujuan}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{perdamaian}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{ga}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{asli}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{ngomong}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{gini}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{a****g}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{katanya}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{harus}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{deportasi}|\text{Negatif}) = \frac{1}{19} = 0,053$
- $P(\text{ya}|\text{Negatif}) = \frac{1}{19} = 0,053$

Netral:

- $P(\text{iri}|\text{Netral}) = \frac{1}{6} = 0,167$
- $P(\text{sama}|\text{Netral}) = \frac{2}{6} = 0,333$
- $P(\text{ketemu}|\text{Netral}) = \frac{1}{6} = 0,167$
- $P(\text{langsung}|\text{Netral}) = \frac{1}{6} = 0,167$
- $P(\text{paus}|\text{Netral}) = \frac{1}{6} = 0,167$
- $P(\text{fransiskus}|\text{Netral}) = \frac{1}{6} = 0,167$

Langkah berikutnya, melakukan perhitungan nilai posterior tertinggi untuk masing-masing kelas sentimen yang ada. Dengan perhitungan ini, *Naïve Bayes* akan menentukan prediksi sentimen dari data uji berdasarkan kelas dengan nilai posterior tertinggi.

Positif:

$$\begin{aligned}
 P(\text{Positif}|\text{Tweet}) &= P(\text{Positif}) \times P(\text{perangko}|\text{Positif}) \times \\
 &P(\text{paus}|\text{Positif}) \times P(\text{fransiskus}|\text{Positif}) \times \\
 &P(\text{dimana}|\text{Positif}) \times P(\text{ya}|\text{Positif}) \\
 &= 0,4 \times 0,08 \times 0,08 \\
 &= \mathbf{0,00256}
 \end{aligned}$$

Negatif:

$$\begin{aligned}
 P(\text{Negatif}|\text{Tweet}) &= P(\text{Negatif}) \times P(\text{perangko}|\text{Negatif}) \times \\
 &P(\text{paus}|\text{Negatif}) \times P(\text{fransiskus}|\text{Negatif}) \times \\
 &P(\text{dimana}|\text{Negatif}) \times P(\text{ya}|\text{Negatif}) \\
 &= 0,4 \times 0,105 \times 0,105 \times 0,053 \\
 &= \mathbf{0,00023327}
 \end{aligned}$$

Netral:

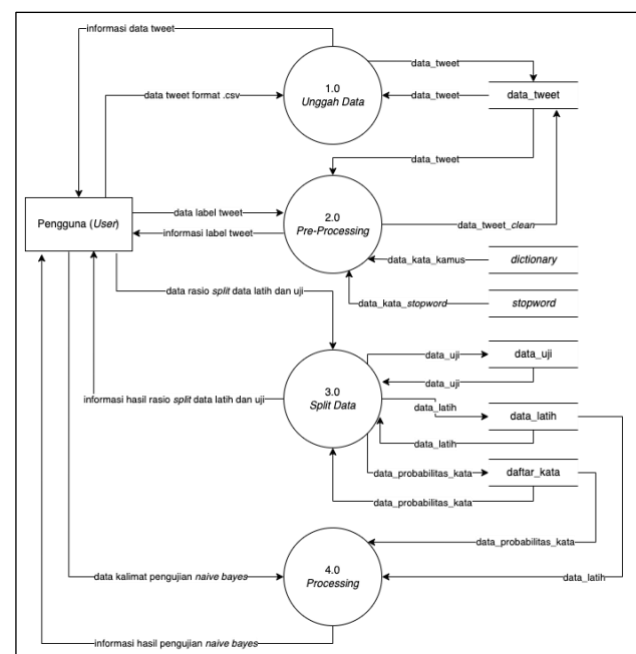
$$\begin{aligned}
 P(\text{Netral}|\text{Tweet}) &= P(\text{Netral}) \times P(\text{perangko}|\text{Netral}) \times \\
 &P(\text{paus}|\text{Netral}) \times P(\text{fransiskus}|\text{Netral}) \times \\
 &P(\text{dimana}|\text{Netral}) \times P(\text{ya}|\text{Netral}) \\
 &= 0,2 \times 0,167 \times 0,167
 \end{aligned}$$

$$= \mathbf{0,005555556}$$

Berdasarkan hasil perhitungan, dapat disimpulkan bahwa tweet “perangko paus fransiskus dimana ya” pada data uji diklasifikasikan sebagai kelas “**Netral**” dengan probabilitas posterior tertinggi, yaitu **0,005555556**.

B. Rancangan

Komponen ini disertakan dalam bentuk *Data Flow Diagram* (DFD) level 1. DFD level 1 memberikan gambaran tentang jalannya program serta alur proses yang terjadi. DFD level 1 menguraikan setiap proses secara rinci, sehingga setiap aliran data terhubung dengan entitas yang terkait. Aplikasi ini memiliki 4 proses yang dilakukan, yaitu mengunggah data, *pre-processing*, *split* data, dan *processing*. Gambar 4 menunjukkan DFD level 1.



Gbr. 4 Data Flow Diagram Level 1

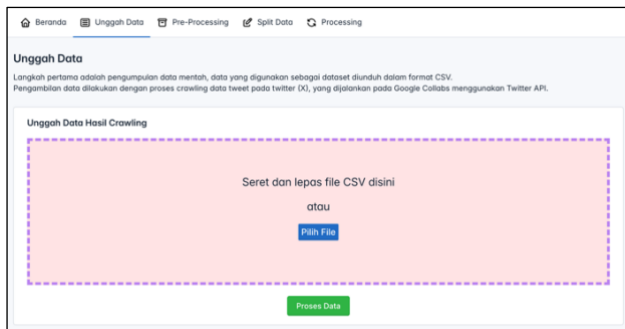
C. Implementasi

Bagian ini menguraikan implementasi antarmuka aplikasi analisis sentimen opini publik terhadap kunjungan Paus Fransiskus di Indonesia menggunakan algoritma *Naïve Bayes*.



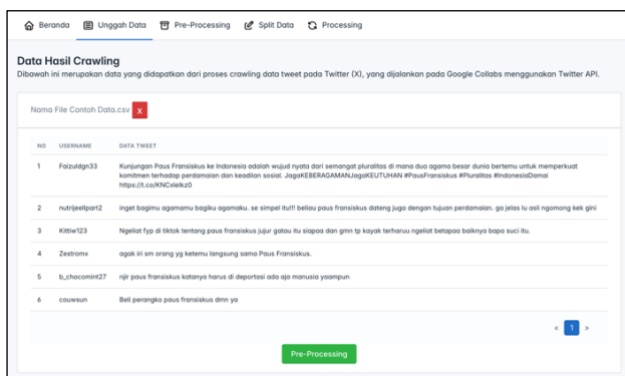
Gbr. 5 Halaman Beranda

Gambar 5 merupakan tampilan utama/beranda dari aplikasi. Pada halaman ini hanya menampilkan nama aplikasi dan gambar Paus Fransiskus.



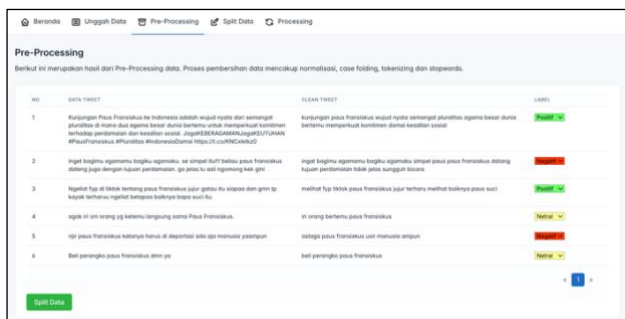
Gbr. 6 Halaman Unggah Data

Gambar 6 merupakan tampilan halaman unggah data. Halaman ini menyediakan opsi bagi pengguna agar melakukan unggah data/file hasil *crawling* dengan format CSV. Terdapat juga tombol proses data yang bisa ditekan untuk melihat isi dari data yang diunggah.



Gbr. 7 Halaman Unggah Data (Hasil)

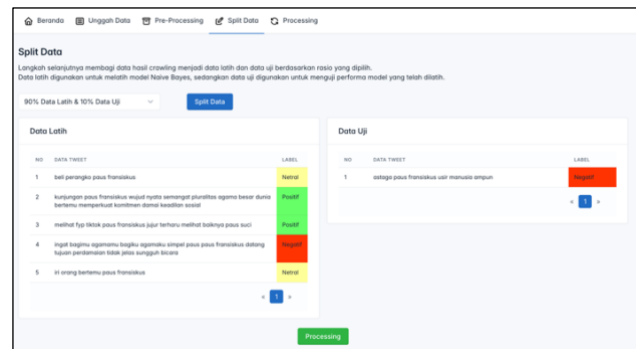
Gambar 7 merupakan tampilan halaman unggah data (hasil). Halaman ini akan menampilkan isi dari data CSV yang telah diunggah sebelumnya. Terdapat juga tombol *pre-processing* yang bisa ditekan untuk melakukan proses ke bagian *pre-processing*.



Gbr. 8 Halaman Pre-Processing

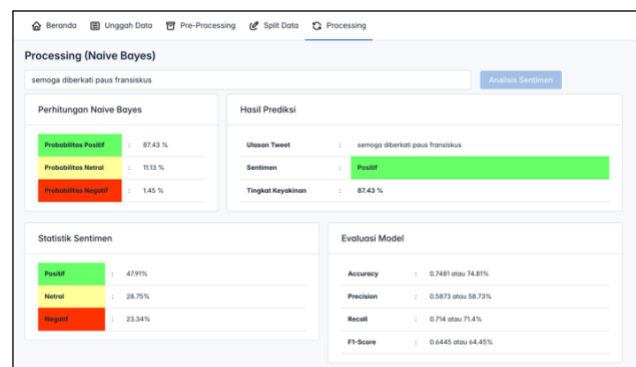
Gambar 8 merupakan tampilan halaman *pre-processing*. Pada halaman ini akan dilakukan proses untuk membersihkan data *tweet* dari hasil *crawling* dengan menggunakan beberapa teknik pemrosesan teks yaitu

normalisasi, tokenisasi, *stopword*, dan *stemming*. Terdapat juga tombol *split data* yang bisa ditekan untuk melakukan proses ke bagian *split data*.



Gbr. 9 Halaman Split Data

Gambar 9 merupakan tampilan halaman *split data*. Pada halaman ini akan dilakukan proses untuk pemisahan *dataset* menggunakan teknik pembagian data dengan memilih persentase rasio data latih dan data uji, yaitu 60:40, 70:30, 80:20, dan 90:10.



Gbr. 10 Halaman Processing

Gambar 10 merupakan tampilan halaman *processing* yang mencakup proses *modelling* dan analisis sentimen untuk pengujian tunggal. Proses *modelling* yang ada pada halaman ini mencakup ekstraksi fitur, pelatihan dan prediksi model serta perhitungan evaluasi metrik. Sedangkan proses analisis sentimen pengujian tunggal pada aplikasi ini merupakan fitur tambahan yang digunakan untuk melihat kinerja model *Naive Bayes* yang telah dilatih, serta memeriksa apakah sentimen berdasarkan data baru sudah sesuai dengan hasilnya.

D. Pengujian

Hasil dari *split data* dari total 448 *dataset* dengan rasio 60:40, 70:30, 80:20, dan 90:10 menghasilkan masing-masing 269 data latih dan 179 data uji, 314 data latih dan 134 data uji, 358 data latih dan 90 data uji, serta 403 data latih dan 45 data uji. Nilai rata-rata dari hasil pengujian berdasarkan 20 kali percobaan dapat dilihat pada Tabel 4.

TABEL IV. PERBANDINGAN HASIL METRIK RATA-RATA

<i>Split</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
60:40	66,44%	49,84%	49,55%	49,70%
70:30	70,91%	54,24%	54,37%	54,30%
80:20	73,26%	58,02%	57,26%	57,61%
90:10	74,37%	59,72%	59,29%	59,48%

Pengujian ini bertujuan untuk menganalisis variasi serta pola konsistensi dalam hasil pengukuran. Hal ini terjadi karena setiap pengujian data diacak sebelum dibagi sesuai rasio perbandingan dan data yang tidak seimbang. Berdasarkan hasil *processing* dari *dataset* sebanyak 448 data, diperoleh hasil rata-rata metrik akurasi, yaitu rasio 90:10 mencapai 74,37%, kemudian rasio 80:20 dengan 73,26%, rasio 70:30 dengan 70,91% dan rasio 60:40 dengan 66,44%. Metrik presisi, *recall*, dan *F1-score* terbaik diperoleh dari rasio 90:10 yaitu 59,72%, 59,29%, dan 59,48%. Evaluasi metrik ini membuktikan bahwa model *Naïve Bayes* mampu bekerja dalam mengklasifikasikan ulasan *tweet*.

V. KESIMPULAN

Berdasarkan pembahasan dan hasil pengujian aplikasi, terdapat beberapa kesimpulan yang dapat diuraikan, yaitu:

1. Dengan penerapan algoritma *Naïve Bayes*, aplikasi analisis sentimen berhasil dikembangkan untuk mengolah opini publik terhadap kunjungan Paus Fransiskus di Indonesia berdasarkan data *tweet* yang diperoleh dari media sosial X.
2. Dari pengujian yang dilakukan, aplikasi terbukti mampu menganalisis ulasan baru dari pengujian tunggal yang dimasukkan, yang kemudian mengklasifikasikan sentimen ke dalam label positif, negatif, atau netral.
3. Algoritma *Naïve Bayes* yang diterapkan dalam aplikasi ini dapat membantu menganalisis opini publik terhadap kunjungan Paus Fransiskus di Indonesia, yang memberikan pemahaman lebih dalam tentang sentimen yang berkembang di masyarakat.

DAFTAR PUSTAKA

- [1] R. Vindua and A. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 dari Media Sosial Twitter Menggunakan Python," *Jurnal Riset Komputer*, vol. 10, no. 2, pp. 479-486, 2023.
- [2] R. Merdiansah, Siska and A. Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 7, no. 1, pp. 221-228, 2024.
- [3] Wikipedia, "Kunjungan Paus Fransiskus ke Indonesia Tahun 2024," 5 Oktober 2024. [Online]. Available: https://id.wikipedia.org/wiki/Kunjungan_Paus_Fransiskus_ke_Indonesia_tahun_2024. [Accessed 9 Oktober 2024].
- [4] A. Nurian and B. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naice Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 829-835, 2023.
- [5] H. Syaucie, A. Nunka and M. Fahrozi, "Analisis Sentimen Ulasan Pengguna Aplikasi Game Sky Children Of The Light Menggunakan Metode Algoritma Naive Bayes (Studi kasus tingkat kepuasan pengguna)," *Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 4, pp. 65-75, 2024.
- [6] P. P. Surya and B. Subbulakshmi, "Sentimental Analysis Using Naive Bayes Classifier," in *The International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 2019.
- [7] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm of the Data Crawler: Twitter," in *The 4th International Conference on Informatics and Computing (ICIC)*, 2019.
- [8] A. Rahayu and A. Rahmat, "Komparasi Algoritma Naive Bayes dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, pp. 349-354, 2022.
- [9] F. Pattiha and Hendry, "Perbandingan Metode K-NN, Naive Bayes, Decision Tree untuk analisis sentimen tweet twitter terkait opini terhadap PT PAL Indonesia," *Jurnal Riset Komputer (JURIKOM)*, vol. 9, no. 2, pp. 506-514, 2022.
- [10] Meliyawati and F. Hasan, "Analisis Sentimen Penggunaan Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naive Bayes," *Kajian ilmiah informatika dan komputer*, vol. 4, no. 4, pp. 2272-2280, 2024.
- [11] R. Abidin and A. Herawati, "Analisis Sentimen Publik Terhadap Kebijakan Program Tabungan Perumahan Rakyat (TAPERA)," *Journal of Information System and Computer*, vol. 4, no. 1, pp. 13-19, 2024.
- [12] H. Faisal, A. Febriandirzah and F. Hasan, "Analisis Sentimen terkait ulasan pada aplikasi PLN Mobile Menggunakan Metode Support Vector Machine," *Jurnal Penerapan Sistem Informasi*, vol. 5, no. 1, pp. 303-312, 2024.
- [13] H. Syaucie, A. Nunka and M. Fahrozi, "Analisis Sentimen Ulasan Pengguna Aplikasi Game Sky Children Of The Light Menggunakan Metode Algoritma Naive Bayes (Studi kasus tingkat kepuasan pengguna)," *Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 4, pp. 65-75, 2024.
- [14] F. Sholekhah, A. Putri and R. Efrizoni, "Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbors Untuk Klasifikasi Metabolik Sindrom," *Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 507-514, 2024.
- [15] A. Averina, H. Hadi and J. Siswantoro, "Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik,"

- TEKNIKA Jurnal Teknologi Informasi dan Komunikasi*, vol. 11, no. 22, pp. 123-128, 2022.
- [16] I. W. Saputro and B. W. Sari, "Naïve Bayes Algorithm Performance Test for Student Study Prediction," *Creative Information Technology Journal*, vol. 6, no. 1, pp. 1-11, 2019.
- [17] S. Tilley and H. Rosenblatt, *System Analysis and Design*, Boston: Cengage Learning, 2016.