

# SISTEM DETEKSI TINGKATAN BAHASA JAWA MENGGUNAKAN METODE NAIVE BAYES

Esti Mulyani<sup>1</sup>, Dea Zulfaniar<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Politeknik Negeri Indramayu

<sup>1</sup>estimulyani@polindra.ac.id, <sup>2</sup>deazulfa22@student.polindra.ac.id

## Abstrak

Bahasa Jawa memiliki sistem unggah-ungguh atau tingkatan bahasa yang terdiri atas Ngoko, Madya, dan Krama Inggil. Setiap tingkatan digunakan dalam konteks sosial yang berbeda, sehingga pemilihan tingkat tutur yang tepat menjadi hal yang penting dalam komunikasi sehari-hari. Namun, banyak penutur, khususnya generasi muda, masih mengalami kesulitan dalam membedakan serta menggunakan ketiga tingkatan tersebut dengan benar. Kesulitan ini muncul karena semakin berkurangnya penggunaan Bahasa Jawa dalam lingkungan keluarga maupun pendidikan formal, serta minimnya media pembelajaran yang praktis dan mudah diakses. Untuk mengatasi permasalahan tersebut, penelitian ini mengembangkan sebuah sistem deteksi dan klasifikasi tingkat tutur Bahasa Jawa menggunakan pendekatan N-Gram sebagai metode ekstraksi fitur dan algoritma Naïve Bayes Classifier sebagai proses klasifikasi utama. Data pelatihan yang digunakan terdiri dari kumpulan teks berbahasa Jawa yang telah diberi label sesuai tingkatan bahasanya. Hasil evaluasi menunjukkan bahwa model mampu mencapai akurasi sebesar 83%. Kinerja terbaik diperoleh pada kelas Ngoko, disusul oleh kelas Madya yang tetap menunjukkan performa cukup stabil. Sementara itu, performa pada kelas Krama Inggil masih relatif rendah akibat ketidakseimbangan data serta kompleksitas variasi kosakata. Meskipun demikian, sistem ini tetap mampu memberikan dukungan pembelajaran yang efektif bagi pengguna dalam memahami tingkatan Bahasa Jawa secara lebih mudah dan terstruktur. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam upaya pelestarian budaya, penguatan literasi bahasa daerah, serta penerapan teknologi sebagai media pembelajaran modern.

**Kata kunci:** bahasa jawa, klasifikasi, n-gram, naïve bayes, TF-IDF, tingkatan bahasa

## 1. Pendahuluan

Indonesia merupakan negara yang memiliki keragaman suku dan budaya. Setiap suku telah menempati wilayah tertentu dan mengembangkan kebudayaannya selama ratusan tahun, yang kemudian berkontribusi pada terbentuknya budaya nasional. Budaya nasional sendiri merupakan hasil perpaduan berbagai unsur kebudayaan di Indonesia yang berkembang dalam masyarakat. Ciri-cirinya tercermin melalui bahasa, seni, tradisi, adat istiadat, serta nilai-nilai kepercayaan yang dianut oleh masing-masing kelompok. Salah satu contohnya adalah kebudayaan yang berbahasa Jawa. Bahasa Jawa adalah salah satu bahasa daerah yang digunakan oleh penduduk terutama di Jawa Timur, Jawa Tengah, dan Yogyakarta. Bahasa Jawa sangat vital dalam kehidupan masyarakat Jawa karena mengandung nilai-nilai luhur budaya mereka.(Angeline, et al 2022).

Bahasa jawa memiliki sebuah karakter yang khas yang tidak dimiliki bahasa lain pada Bahasa Jawa memiliki sebuah tingkatan bahasa yang tidak mudah untuk dipahami. Tingkatan ini adalah Ngoko, Madya, dan Krama (Puji Arfianingrum, 2020). Ngoko adalah tingkatan yang digunakan kepada seseorang dengan derajat atau tingkatan lebih rendah contohnya adik atau orang yang lebih muda.

Sedangkan madya dipakai kepada seseorang yang memiliki tingkatan yang sederajat atau sama. Dan Krama merupakan tingkatan yang digunakan pada seseorang yang lebih tua dan juga dapat digunakan pada seseorang yang memiliki derajat lebih tinggi (Dewi, et al 2025).

Seiring dengan berkembangnya teknologi informasi, yang semakin pesat dan tidak dikendalikan (Gita Segara & Irwan Padli Nasution, 2025) maka dapat dilakukan pelestarian Bahasa Jawa melalui media teknologi. Di sisi lain, metode klasifikasi teks menggunakan metode Naïve Bayes terbukti efektif untuk melakukan analisis dan klasifikasi teks (Putri et al., 2025). Namun, pada penelitian mengenai penerapan klasifikasi teks bagi tingkatan Bahasa Jawa masih tergolong sedikit, sehingga penelitian ini mencoba mengisi kekosongan tersebut.

Seiring berkembangnya teknologi, berbagai penelitian telah mengkaji upaya pelestarian bahasa Jawa melalui aplikasi pembelajaran dan sistem berbasis digital. Misalnya, (Hidayatullah, et al 2023) mengembangkan model *word-level language identification* untuk mendeteksi bahasa Jawa dalam teks campuran menggunakan pendekatan pembelajaran mesin modern. Selain itu, metode klasifikasi berbasis machine learning seperti Naïve Bayes, SVM, dan KNN telah banyak dieksplorasi

pada teks berbahasa Indonesia maupun Jawa untuk tugas klasifikasi dan analisis sentimen (Nurhidayat & Dewi, 2023) (Garcia dkk., n.d.) (S. D. A. Putri, et al 2021). Dibandingkan dengan SVM dan KNN, metode Naïve Bayes memiliki keunggulan berupa proses pelatihan yang lebih ringan karena hanya memerlukan perhitungan probabilitas, sehingga dapat bekerja secara efisien pada data berdimensi tinggi seperti teks. Selain itu, Naïve Bayes juga tetap menunjukkan kinerja yang baik meskipun dataset yang digunakan tidak terlalu besar.

Penelitian mengenai Abusive Language Detection berbasis Naïve Bayes dan SVM juga telah diuji pada teks Jawa dan Sunda dengan hasil yang cukup baik (Pradipta et al., 2021). Temuan-temuan tersebut memperkuat bahwa teknologi pemrosesan bahasa alami (NLP) memiliki potensi besar dalam mendukung pelestarian bahasa daerah melalui penerapan metode klasifikasi yang tepat.

Selain itu, pentingnya klasifikasi tingkatan bahasa Jawa juga telah dibahas dalam beberapa penelitian terdahulu. Ardhana et al. (2019) menggunakan metode Multinomial Naïve Bayes dan N-Gram untuk mengklasifikasikan tingkat bahasa Jawa pada teks artikel, dan hasilnya menunjukkan bahwa pendekatan tersebut efektif untuk mengenali variasi tingkatan bahasa. Angeline et al. (2022) juga melakukan penelitian klasifikasi dialek bahasa Jawa menggunakan Naïve Bayes dengan hasil akurasi yang baik. Namun demikian, penelitian terkait masih sangat terbatas, khususnya yang fokus pada deteksi otomatis tiga tingkatan utama bahasa Jawa (Ngoko, Madya, dan Krama Ingil) dengan kombinasi metode N-Gram dan Naïve Bayes yang diimplementasikan dalam sistem berbasis web interaktif.

Berdasarkan latar belakang tersebut, penelitian ini difokuskan pada pengembangan sistem deteksi dan klasifikasi tingkatan bahasa Jawa menggunakan metode N-Gram dan algoritma Naïve Bayes. Kontribusi penelitian ini terletak pada penyediaan media pembelajaran interaktif yang dapat membantu masyarakat memahami perbedaan tingkatan bahasa Jawa, sekaligus mendukung pelestarian bahasa daerah melalui teknologi. Adapun hipotesis penelitian ini adalah bahwa kombinasi metode N-Gram dan Naïve Bayes mampu mencapai akurasi yang baik ( $\geq 80\%$ ) dalam melakukan klasifikasi tingkatan bahasa Jawa.

## 2. Metode

### 2.1. Pengumpulan Data

Pengumpulan data dilakukan melalui studi literatur dari buku, dokumen tertulis, serta sumber terbuka (open source) seperti GitHub dan dokumen daring lainnya. Dataset yang digunakan dalam pelatihan model terdiri dari 1.000 kalimat berbahasa Jawa tanpa label yang diambil dari sumber terbuka tersebut. Setelah data terkumpul, proses pelabelan

dilakukan secara otomatis dengan merujuk pada kamus tingkatan bahasa Jawa. Setiap kalimat dianalisis berdasarkan jumlah kata yang termasuk dalam kategori ngoko, madya, atau krama ingil; label kemudian ditentukan dari kategori yang paling dominan dalam kalimat tersebut.

Meskipun metode pelabelan otomatis ini efisien, terdapat keterbatasan karena tidak mempertimbangkan konteks secara menyeluruh. Beberapa kata dapat digunakan dalam dua tingkatan bahasa berbeda tergantung konteks penggunaannya, sehingga berpotensi menimbulkan kesalahan label. Untuk meminimalkan risiko tersebut, dilakukan proses Quality Control (QC) melalui teknik sampling manual terarah. Sejumlah data diperiksa ulang oleh penutur asli secara acak namun difokuskan pada kelas minoritas yang lebih rentan salah label. Langkah ini memastikan kualitas dataset tetap terjaga sehingga data latih yang digunakan lebih andal dan mendukung peningkatan akurasi model klasifikasi. Tabel 1 menunjukkan statistik pelabelan di setiap kelasnya.

Tabel 1 data statistik pelabelan setiap kelas

Kelas	Jumlah Data	Persentase
Ngoko	831	83,1 %
Madya	102	10,2 %
Krama Ingil	67	6,7 %

Pada Tabel 2 menunjukkan contoh dataset yang dipilih untuk penelitian ini.

Tabel 2. Dataset Tanpa *Preprocessing*

Kalimat	Label
Bocah kuwi seneng nggambarnya sesawangan sing asri.	Ngoko
Dheweke mlaku kanthi alon-alon.	Krama ingil
Simin ngampeng kancane.	Madya

### 2.2. Pre Processing

*Preprocessing* merupakan proses yang digunakan untuk menormalkan istilah – istilah yang berasal dari kalimat. Hal ini dilakukan untuk menerima data latih yang baik serta fitur – fitur yang diekstrak nantinya sinkron dengan yang diinginkan sehingga mampu menyederhanakan pemrosesan data. (SLAM et al., 2025).

Pada mengambil data melalui studi literatur ini masih data mentah, dalam penelitian model deteksi dan klasifikasi ini membutuhkan data yang tidak mengandung elemen yang tidak relevan atau mengganggu. Oleh karena itu dilakukan tahap *preprocessing* untuk meningkatkan kualitas data. Tahap ini juga bertujuan untuk menyederhanakan representasi teks, mengurangi noise, dan menyamakan format sehingga dapat meningkatkan akurasi algoritma klasifikasi.

Berikut beberapa tahapan pemrosesan data yang digunakan dalam penelitian ini:

**Lowering Case**, merupakan teknik *preprocessing* data untuk mengubah huruf besar ke huruf kecil

secara keseluruhan pada dataset (Cahyo Prianto et al., 2023). Contoh penerapan lowering case:  
Sebelum: Aku Arep MANGAN  
Sesudah: aku arep mangan

**Cleansing**, yaitu proses yang bertujuan untuk mengurangi noise pada teks dataset, seperti karakter dan elemen yang tidak berguna. Dalam Cleaning juga dilakukan penghapusan baris yang duplikat (Sentimen et al., 2025). Contoh penerapan cleansing:  
Sebelum: aku badhé dhahar.):  
Sesudah: aku badhe dhahar

**Tokenisasi**, merupakan pemecahan sebuah kalimat menjadi token atau kata individual (Nauli et al., 2025). Contoh penerapan tokenisasi:

Sebelum: aku arep mangan  
Sesudah: aku, arep, mangan

Contoh transformasi pada salah satu label tingkatan ditunjukkan pada Tabel 3.

Tabel 3 Dataset dengan *Preprocessing*

Kalimat			Label
bocah	kuwi	seneng	nggambar
sesawangan	sing asri		Ngoko
dheweke	mlaku	kanthi	Krama
simin	alonalon		Inggil
ngampeling	kancane		Madya

### 2.3. TF-IDF

TF atau Frekuensi Term mengacu pada seberapa sering sebuah kata muncul dalam sebuah teks, sedangkan IDF adalah indeks frekuensi dokumen yang dibalik. Konsep utama di balik TF-IDF adalah bahwa kata yang muncul lebih sering dalam dokumen tertentu dan jarang ditemukan di dokumen lain harus dianggap lebih penting karena lebih bermanfaat untuk pengklasifikasian (Arasy & Agustian, 2025).

Setelah data terkumpul dan melalui tahap *preprocessing*, selanjutnya adalah melakukan ekstraksi fitur yang berfungsi untuk mengubah data teks menjadi bentuk angka agar bisa digunakan oleh algoritma klasifikasi. Pada penelitian ini, metode yang digunakan adalah *Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan untuk mengekstraksi fitur dari dataset yang telah melalui tahap *preprocessing*.

Metode TF-IDF adalah metode yang digunakan untuk menghitung bobot setiap kata dalam sebuah dokumen. Metode ini dikenal efisien, mudah digunakan, dan menghasilkan keakuratan yang baik. Metode ini menghitung bobot setiap token  $t$  dalam dokumen  $d$  dengan menggunakan Persamaan 1.

$$W_{dt} = tf_{dt} * IDF_t \quad (1)$$

Dimana:

- $d$  : dokumen ke- $d$
- $t$  : kata ke- $t$  dari kata kunci
- $W$  : bobot dokumen ke- $d$  terhadap kata ke- $t$
- $tf$  : banyaknya kata yang dicari pada sebuah

$IDF$	dokumen
	: Inversed document Frequency Nilai
$IDF$	IDF didapatkan dari
$IDF$	: $\log_2 (D/df)$
$D$	: total dokumen
$df$	: banyak dokumen yang mengandung kata yang dicari

Setelah mengetahui bobot (W) masing-masing dokumen, selanjutnya dilakukan pengurutan sehingga dokumen dengan nilai W yang lebih besar memiliki tingkat kemiripan yang lebih tinggi terhadap kata kunci, sebaliknya dokumen dengan nilai W yang lebih kecil memiliki tingkat kemiripan yang lebih rendah.

### 2.4. N-Gram

N-Gram adalah sejumlah pecahan kata yang dihasilkan dari sebuah kalimat. Metode N-Gram bisa juga digunakan untuk membangkitkan kata atau karakter. N-Gram dapat dibentuk sedemikian rupa berdasarkan dari kata-kata sebelumnya dan berikutnya (Nurhidayat & Dewi, 2023). Dalam penelitian ini, metode N-Gram digunakan sebagai teknik ekstraksi fitur untuk memecah kalimat berbahasa Jawa menjadi unit-unit kata yang dapat dianalisis oleh sistem. N-Gram bekerja dengan cara membagi teks menjadi urutan kata berdasarkan jumlah kata yang berurutan dalam kalimat, sehingga membantu sistem dalam menangkap konteks dan pola penggunaan kata dalam setiap tingkatan bahasa Jawa.

Penelitian ini menggunakan model Unigram sebagai representasi teks. Unigram merupakan bentuk sederhana dari N-Gram dengan  $n = 1$ , di mana kalimat dipecah menjadi kata-kata tunggal. Setiap kata diperlakukan sebagai fitur independen yang berdiri sendiri dalam proses klasifikasi. Penerapan Unigram sebagai ekstraksi fitur karena bahasa jawa memiliki perbedaan tingkatan yang lebih banyak ditentukan oleh variasi kosakata daripada struktur kalimat (Zaman et al., 2015).

### 2.5. SMOTE

SMOTE adalah teknik oversampling yang bekerja dengan membuat data sintetis untuk kelas yang jumlahnya sedikit. Berbeda dari oversampling biasa yang hanya melakukan penggandaan data (copy-paste), SMOTE membuat contoh baru dengan cara menghitung interpolasi antara data minoritas yang sudah ada. Pada penelitian kali ini SMOTE digunakan untuk menyeimbangkan dataset yang timpang antar kelasnya.

### 2.6. Naïve Bayes

Naive Bayes sering digunakan untuk mengklasifikasikan teks dalam berbagai bahasa. Algoritma Naive Bayes merupakan metode klasifikasi yang memanfaatkan probabilitas dan

statistik berdasarkan *Teorema Bayes*. Salah satu keunggulan algoritma ini adalah kemampuannya memberikan akurasi yang baik meskipun perhitungannya sederhana (Rayuwati et al., 2022).

Penelitian ini menggunakan Naïve Bayes untuk melakukan klasifikasi tingkatan Bahasa jawa. Dengan menggunakan probabilitas dan statistik yang berdasarkan pada *teorema bayes*. Kelebihan menggunakan Naïve Bayes memiliki akurasi yang tinggi meskipun perhitungannya sederhana.

Naïve Bayes memiliki beberapa jenis, salah satunya adalah Multinomial Naïve Bayes. Algoritma Multinomial Naïve Bayes merupakan metode pembelajaran berbasis probabilitas yang menggunakan *teorema Bayes*. Algoritma ini sering digunakan dalam *Natural Language Processing* (NLP). Cara kerja algoritma ini berdasarkan konsep frekuensi kata, yaitu seberapa sering suatu kata muncul dalam satu dokumen. Model ini menjelaskan dua hal, yaitu apakah suatu kata muncul dalam dokumen dan seberapa sering kata tersebut muncul. Multinomial Naïve Bayes dapat diformulasikan menjadi Persamaan 2.

$$P(p|n) \propto P(p) \prod_{1 \leq k \leq n} P(t_k|p) \quad (2)$$

dimana  $P(t_k|p)$ : probabilitas munculnya dokumen teks ( $t_k$ ),  $n$  adalah munculnya jumlah dokumen dan  $p$  adalah polaritas. Kemudian untuk menghitung polaritasnya atau dokumen yang mempunyai kemiripan dirumuskan pada Persamaan 3.

$$P(t_k|p) = \frac{\text{count}(t_k|p) + 1}{\text{count}(t_p) + |V|} \quad (3)$$

Di mana  $(t_k | p)$  adalah jumlah  $t_k$  yang muncul di dokumen teks dengan polaritas  $p$ , dan jumlah  $(tp)$  berarti jumlah token yang terdapat dalam dataset polaritas  $p$ .

## 2.7. Evaluasi

Evaluasi bertujuan untuk menilai hasil uji coba sistem yang dibuat apakah telah sesuai antara hasil sistem analisis sentimen dengan hasil sebenarnya. *Confusion matrix* adalah tabel yang digunakan untuk menganalisis seberapa baik keakuratan sebuah metode klasifikasi untuk memprediksi kelas suatu data (Gifari et al., 2022).

Evaluasi ini bertujuan untuk mengetahui seberapa baik model mampu memprediksi label kelas berdasarkan fitur teks yang diberikan. Evaluasi dilakukan setelah model telah dilatih menggunakan dataset yang telah melewati proses *preprocessing*, TF-IDF, dan Model Naïve Bayes. Dengan demikian, performa serta keandalan model dapat diketahui sebelum diaplikasikan dalam penggunaan nyata.

Dalam penelitian ini, beberapa metrik yang digunakan untuk mengevaluasi model meliputi *confusion matrix*, *accuracy*, *precision*, *recall*, dan

*F1-Score*. Setiap metrik tersebut memiliki peran tersendiri dalam menampilkan gambaran mengenai seberapa baik model berkinerja dari berbagai sudut pandang, sehingga memastikan evaluasi dilakukan secara menyeluruh terhadap hasil klasifikasi yang dihasilkan oleh model. Berikut penjelasan masing-masing metrik evaluasi yang digunakan dalam penelitian ini:

**Confusion Matrix**, dalam hal ini berisikan :

**True Positive (TP)**: Prediksi positif, dan benar positif

**True Negative (TN)**: Prediksi negatif, dan benar negatif

**False Positive (FP)**: Prediksi positif, padahal sebenarnya negatif

**False Negative (FN)**: Prediksi negatif, padahal sebenarnya positif.

**Accuracy**, untuk mengukur proporsi prediksi benar dibandingkan seluruh prediksi, dengan formula ditunjukkan pada Persamaan 4.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

**Precision**, untuk mengetahui seberapa tepat prediksi positif yang berhasil diprediksi, ditunjukkan pada Persamaan 5.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

**Recall**, untuk menilai kemampuan model dalam mengidentifikasi semua data yang benar-benar positif. Formula perhitungan recall ditunjukkan Persamaan 6.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

F1-Score, seperti ditunjukkan pada Persamaan 7, merupakan rata-rata dari *precision* dan *recall*.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

## 2.8. Implementasi Sistem

Pada implementasi sistem ini bertujuan untuk merealisasikan seluruh tahapan proses klasifikasi tingkatan Bahasa jawa kedalam sebuah aplikasi berbasis website yang dapat digunakan secara interaktif oleh pengguna. Sistem ini dikembangkan menggunakan *framework* Flask dikarenakan Flask ini ringan, dan mudah diintegrasikan dengan pustaka pemrosesan bahasa alami dan pembelajaran mesin seperti scikit-learn. Aplikasi ini dirancang untuk menerima input beberapa kalimat Bahasa Jawa, memprosesnya, lalu menampilkan hasil klasifikasi secara real-time.

Sistem ini menggunakan metode N-Gram untuk proses ekstraksi fitur dan algoritma Naive Bayes untuk proses klasifikasi. Dengan penerapan sistem ini, kalimat berbahasa Jawa yang diinputkan akan diproses melalui tahap *preprocessing*, ekstraksi fitur, hingga klasifikasi sehingga dapat mengidentifikasi tingkatan Bahasa Jawa, yakni Ngoko, Madya, dan Krama inggil, dengan lebih cepat dan akurat. Implementasi sistem ini diharapkan dapat membantu dalam upaya pelestarian Bahasa Jawa serta memberikan pemahaman kepada masyarakat mengenai penggunaan tingkatan Bahasa Jawa secara tepat sesuai dengan konteksnya.

## 2.9. Alur Kerja Sistem

Alur penelitian pada sistem deteksi dan klasifikasi tingkatan Bahasa Jawa menggunakan metode N-Gram dan algoritma *Naive Bayes* dapat dijelaskan sebagai berikut:

Penelitian diawali dengan tahap pengambilan dataset, di mana kalimat-kalimat berbahasa Jawa dikumpulkan dari berbagai sumber, seperti media sosial, artikel daring, dan sumber terbuka lainnya. Setelah data terkumpul, dilakukan labeling dataset dengan memberikan label tingkatan Bahasa Jawa pada setiap kalimat, yaitu ngoko, madya, atau krama inggil. Dataset yang telah dilabeli kemudian digabungkan dalam satu file agar memudahkan proses pengolahan data pada tahap berikutnya.

Tahap selanjutnya adalah *preprocessing* data, yang meliputi penghapusan *stopword* (kata yang tidak memiliki makna penting), *tokenization* (memecah kalimat menjadi kata), serta *noise removal* untuk menghapus karakter tidak penting seperti tanda baca, angka, dan simbol.

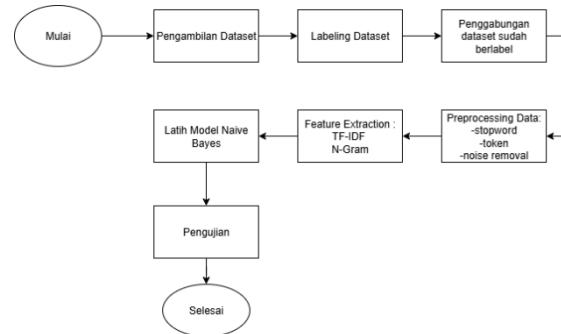
Setelah tahap *preprocessing* selesai, dilakukan feature extraction dengan menggunakan metode TF-IDF dan N-Gram. TF-IDF digunakan untuk memberikan bobot pada setiap kata sehingga dapat menunjukkan tingkat kepentingan kata dalam kalimat, sedangkan N-Gram digunakan untuk menangkap urutan kata (unigram dan bigram) sehingga sistem dapat memahami pola kata yang sering muncul secara berurutan dalam setiap tingkatan Bahasa Jawa.

Tahap berikutnya adalah pelatihan model menggunakan algoritma *Naive Bayes* dengan memanfaatkan data yang sudah diekstraksi fiturnya. Pada tahap ini, model akan mempelajari pola penggunaan kata pada masing-masing tingkatan Bahasa Jawa untuk digunakan pada tahap klasifikasi.

Kemudian jika proses pelatihan telah selesai, maka selanjutnya adalah pengujian pada model menggunakan data uji untuk mengevaluasi performa sistem dalam mendeteksi dan mengklasifikasikan tingkatan Bahasa Jawa. Penelitian ini diakhiri dengan tahap selesai setelah semua tahapan telah dilaksanakan dan model telah dievaluasi performanya untuk mengetahui tingkat akurasi dan keandalannya

dalam mendeteksi serta mengklasifikasikan tingkatan Bahasa Jawa secara otomatis.

Alur kerja sistem secara keseluruhan terlihat pada Gambar 1.



Gambar 1. Alur Kerja Sistem

Secara umum, tahapan penelitian ini ditunjukkan pada Gambar 1, yang meliputi pengumpulan data, *preprocessing*, ekstraksi fitur, pelatihan model, pengujian, dan evaluasi.

## 3. Hasil dan Pembahasan

### 3.1 Hasil Sistem



Gambar 2. Interface Awal

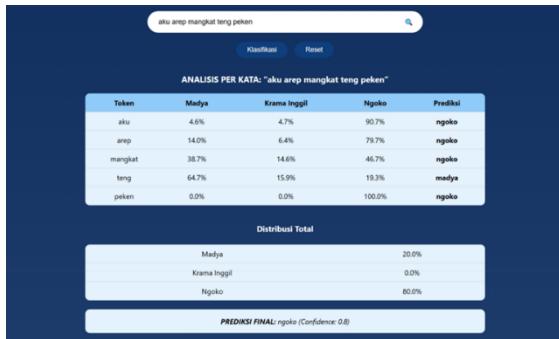
Gambar 2 menampilkan tampilan antarmuka sistem deteksi dan klasifikasi tingkatan Bahasa Jawa pada bagian input kalimat. Antarmuka ini dirancang secara sederhana dan user-friendly agar memudahkan pengguna dalam melakukan input kalimat berbahasa Jawa yang ingin dideteksi tingkatan bahasanya.

Pada bagian atas *form* terdapat deskripsi singkat untuk memberikan pemahaman kepada pengguna mengenai tingkatan Bahasa Jawa yang akan diprediksi oleh sistem, yaitu ngoko, madya, dan krama inggil.

Pada bagian bawah antarmuka, terdapat **kolom input teks** yang digunakan untuk memasukkan kalimat berbahasa Jawa yang ingin diklasifikasikan tingkatan bahasanya oleh sistem. Setelah kalimat dimasukkan, pengguna dapat menekan tombol **“Klasifikasi”** untuk memproses inputan, dan sistem akan memberikan hasil klasifikasi sesuai tingkatan Bahasa Jawa yang terdeteksi. Tersedia pula tombol **“Reset”** yang dapat digunakan untuk menghapus inputan jika pengguna ingin memasukkan kalimat lain.

Antarmuka ini hanya berfungsi sebagai media input pengguna, sedangkan proses klasifikasi akan

dilakukan oleh sistem di sisi backend dengan memanfaatkan model klasifikasi yang telah dibangun menggunakan metode N-Gram dan *algoritma Naive Bayes*. Dengan adanya antarmuka ini, pengguna dapat dengan mudah memanfaatkan sistem untuk memeriksa tingkatan Bahasa Jawa dari kalimat yang mereka masukkan secara praktis.



Gambar 3. Interface Hasil Klasifikasi

Gambar 3 menunjukkan hasil klasifikasi kalimat “aku arep mangkat teng peken”. Dari analisis per kata, sebagian besar token seperti *aku*, *arep*, *mangkat*, dan *peken* terdeteksi sebagai Ngoko, sedangkan kata *teng* terdeteksi sebagai Madya. Hasil akumulasi distribusi menunjukkan 80% Ngoko, 20% Madya, dan 0% Krama Inggil. Dengan demikian, sistem memberikan prediksi akhir Ngoko dengan tingkat keyakinan 0,8.

Menariknya, kata *peken* ditampilkan dengan nilai 0% pada tabel distribusi probabilitas, tetapi tetap terprediksi sebagai Ngoko. Hal ini terjadi karena dalam dataset pelatihan, kata *peken* hanya muncul di kelas Ngoko dan tidak ditemukan pada kelas lain. Akibatnya, meskipun nilai probabilitas relatif terhadap Madya dan Krama Inggil adalah nol, sistem secara otomatis menetapkan kata tersebut ke kelas Ngoko.

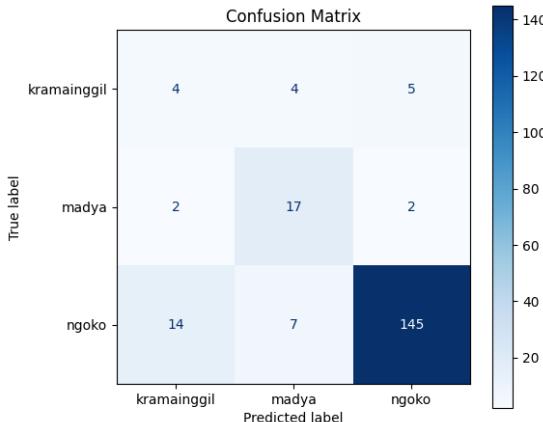
Hasil ini menunjukkan bahwa meskipun ada kata Madya dalam kalimat (*teng*), dominasi kosakata Ngoko membuat sistem mengklasifikasikan kalimat secara keseluruhan sebagai Ngoko.

Melalui tampilan ini, pengguna dapat memahami bagaimana sistem menganalisis setiap kata dalam kalimat yang dimasukkan serta mendapatkan informasi tingkatan Bahasa Jawa secara menyeluruh dari kalimat tersebut. Tampilan hasil prediksi ini juga membantu pengguna untuk mempelajari dan membedakan tingkatan Bahasa Jawa secara praktis dengan bukti distribusi analisis yang transparan.

### 3.2 Hasil Evaluasi

Evaluasi dilakukan untuk mengetahui performa sistem klasifikasi tingkatan Bahasa Jawa menggunakan algoritma *Multinomial Naive Bayes* dengan fitur TF-IDF n-gram (1,2) dan penyeimbangan data menggunakan SMOTE. Dataset yang digunakan terdiri dari tiga kelas tingkatan, yaitu Krama Inggil, Madya, dan Ngoko, dengan proporsi

data yang berbeda antar kelas. Evaluasi dilakukan menggunakan classification report, confusion matrix, dan grafik distribusi prediksi label agar performa dapat terlihat secara menyeluruh. Dijelaskan pada Gambar 4 dan Tabel 4.



Gambar 4. Confusion Matrix

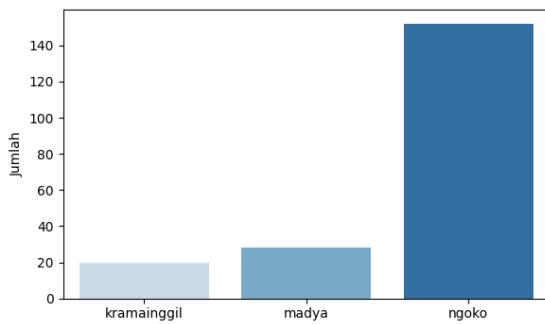
Berdasarkan confusion matrix pada Gambar 4, dihasilkan, terlihat bahwa model memprediksi dengan sangat baik pada kelas Ngoko dengan jumlah prediksi yang benar mencapai 145 dari total 166 data yang tersedia. Namun, masih terdapat kesalahan prediksi ke kelas Krama Inggil dan Madya, meskipun jumlahnya tidak terlalu signifikan. Pada kelas Madya, model mampu memprediksi dengan cukup baik dengan 17 prediksi benar dari 21 data, menunjukkan model dapat mengenali pola kata pada tingkatan Madya dengan baik meskipun masih terdapat beberapa kesalahan prediksi. Sementara itu, pada kelas Krama Inggil, performa model masih tergolong rendah dengan hanya 4 prediksi benar dari 13 data, dan sebagian besar data pada kelas ini cenderung terprediksi menjadi Ngoko.

Tabel 4 Classification Report

Label	Precision	Recall	F1-Score	Support
Kramainggil	0.20	0.31	0.24	13
Madya	0.61	0.81	0.69	21
Ngoko	0.95	0.87	0.91	166
Accuracy			0.83	

Classification report pada Tabel 3 menunjukkan Laporan klasifikasi menunjukkan model memiliki akurasi sebesar 83%. Nilai *precision* pada kelas Ngoko sangat tinggi, yaitu 0.95, sedangkan nilai *recall* mencapai 0.87, yang menandakan model mampu mengenali kata-kata dalam kelas Ngoko dengan sangat baik. Untuk kelas Madya, nilai *precision* adalah 0.61 dan *recall* sebesar 0.81, menunjukkan model cukup baik dalam mengenali kelas ini. Namun, pada kelas Krama Inggil, nilai *precision* hanya 0.20 dan *recall* 0.31, yang menunjukkan performa model di kelas ini masih sangat rendah dibandingkan kelas lainnya. Hal ini menunjukkan adanya ketidakseimbangan dalam distribusi data di dataset, yang berdampak pada kemampuan prediksi model, terutama pada kelas

Krama Inggil yang memiliki jumlah data lebih sedikit dibandingkan kelas lainnya.



Gambar 5. Grafik Distribusi Label

Grafik distribusi prediksi label pada gambar 5 juga memperlihatkan dominasi prediksi pada kelas Ngoko, terlihat dari jumlah prediksi yang sangat tinggi dibandingkan kelas Madya dan Krama Inggil. Hal ini mengindikasikan bahwa model masih cenderung memprediksi kata-kata yang diberikan ke dalam kelas Ngoko karena distribusi data yang lebih banyak pada kelas ini selama pelatihan model, serta karakteristik model *Naïve Bayes* yang sensitif terhadap distribusi data pada saat pelatihan.

Secara keseluruhan, sistem klasifikasi tingkatan kata Bahasa Jawa ini telah mampu melakukan klasifikasi dengan akurasi yang cukup baik dan dapat mengenali tingkatan kata secara otomatis, terutama pada kelas Ngoko dan Madya. Namun, performa model pada kelas Krama Inggil masih perlu ditingkatkan dengan penambahan jumlah data atau penggunaan teknik balancing yang lebih optimal agar distribusi prediksi menjadi lebih seimbang dan akurasi pada semua tingkatan dapat meningkat secara merata. Evaluasi ini menjadi dasar untuk penyempurnaan model pada tahap pengembangan berikutnya agar dapat mendukung deteksi tingkatan kata Bahasa Jawa secara akurat dan seimbang.

### 3.3 Analisis Hasil

Hasil analisis pada sistem klasifikasi tingkatan kata Bahasa Jawa menunjukkan bahwa model berhasil mengklasifikasikan kata-kata ke dalam tiga tingkatan, yaitu Krama Inggil, Madya, dan Ngoko dengan tingkat akurasi yang cukup tinggi. Dalam pengujian dengan data uji, sistem berhasil mencapai akurasi sebesar 83%. Nilai *precision* dan *recall* pada kelas Ngoko juga cukup tinggi, yaitu 0.95 untuk *precision* dan 0.87 untuk *recall*. Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali kata-kata yang berkategori Ngoko.

Pada kelas Madya, nilai *precision* yang diperoleh adalah 0.61 dengan *recall* 0.81, menunjukkan bahwa sistem cukup baik dalam mendeteksi kata dengan tingkatan Madya, meskipun masih terdapat beberapa kesalahan klasifikasi. Sementara itu, pada kelas Krama Inggil, sistem masih memiliki keterbatasan

dengan nilai *precision* sebesar 0.20 dan *recall* sebesar 0.31, yang menampilkan bahwa model masih kesulitan untuk mengenali kata dengan tingkatan ini. Hal ini kemungkinan disebabkan oleh jumlah data yang tidak seimbang pada dataset, di mana kelas Krama Inggil memiliki jumlah data yang jauh lebih sedikit dibandingkan kelas Ngoko.

Hasil confusion matrix menunjukkan bahwa masih terdapat kesalahan prediksi pada kelas Krama Inggil, misalnya terdapat 4 data Krama Inggil yang justru terkласifikasi sebagai Madya. Hal ini dapat disebabkan oleh adanya overlap kosakata antara kedua tingkatan bahasa tersebut. Beberapa kata dalam Krama Inggil juga dipakai pada Madya dalam konteks tertentu, sehingga sistem yang hanya berbasis unigram cenderung kesulitan membedakan secara akurat. Selain itu, jumlah data pada kelas Krama Inggil yang lebih sedikit membuat model kurang mampu mempelajari pola kosakata secara optimal dibandingkan Ngoko atau Madya.

Visualisasi distribusi prediksi menunjukkan dominasi prediksi pada kelas Ngoko, yang menegaskan bahwa sistem cenderung memprediksi kata ke dalam kelas tersebut karena karakteristik data pelatihan yang memiliki jumlah data paling banyak pada kelas ini.

Validasi hasil penelitian ini juga sejalan dengan penelitian lain. Model Multinomial Naïve Bayes pada penelitian ini menghasilkan akurasi 0.83, yang mendukung temuan penelitian (Angeline, et al 2022) bahwa metode Naïve Bayes bekerja baik dalam klasifikasi dialek Jawa dengan akurasi di atas 80%. Hasil ini juga konsisten dengan penelitian (Haris, et al 2025) dan (S. D. A. Putri, et al 2021) yang melaporkan performa Naïve Bayes dalam klasifikasi teks Bahasa Indonesia dengan akurasi rata-rata lebih dari 80%. Dengan demikian, hasil penelitian ini memperkuat bukti bahwa Naïve Bayes merupakan metode yang andal untuk klasifikasi teks berbasis bahasa daerah.

Secara keseluruhan, hasil analisis ini menunjukkan bahwa sistem yang dibangun telah mampu melakukan klasifikasi tingkatan kata Bahasa Jawa secara otomatis dengan performa yang cukup baik pada kelas Ngoko dan Madya, serta masih memerlukan perbaikan pada kelas Krama Inggil untuk meningkatkan keseimbangan prediksi antar kelas agar sistem dapat lebih akurat dan optimal dalam mendeteksi tingkatan kata Bahasa Jawa pada kalimat yang diinputkan oleh pengguna.

### 3.4 Implikasi Hasil

Implikasi dari hasil analisis ini menunjukkan bahwa sistem klasifikasi tingkatan kata Bahasa Jawa yang dibangun dapat membantu pengguna dalam memahami tingkatan Bahasa Jawa pada sebuah kalimat secara otomatis, khususnya pada kata-kata dengan tingkatan Ngoko dan Madya. Dengan akurasi yang cukup tinggi pada kedua tingkatan ini, sistem

dapat digunakan sebagai sarana belajar untuk membantu siswa atau mahasiswa, peneliti bahasa, maupun masyarakat umum dalam mempelajari tingkatan Bahasa Jawa secara praktis dan efisien.

Namun, hasil ini juga mengimplikasikan perlunya pengembangan lebih lanjut untuk meningkatkan akurasi pada kelas Krama Inggil, mengingat penggunaan Krama Inggil penting dalam situasi formal atau komunikasi dengan pihak yang lebih dihormati dalam budaya Jawa. Hal ini menunjukkan bahwa keseimbangan data dalam pelatihan model menjadi aspek yang penting untuk diperhatikan agar sistem dapat bekerja secara optimal pada semua tingkatan bahasa.

Selain dari itu, sistem ini juga dapat digunakan sebagai dasar pengembangan aplikasi pembelajaran Bahasa Jawa yang lebih interaktif, yang memungkinkan pengguna mempraktikkan penggunaan tingkatan Bahasa Jawa secara mandiri dengan umpan balik yang cepat. Implikasi lainnya, sistem ini dapat diintegrasikan ke dalam aplikasi keyboard atau asisten digital yang membantu pengguna memilih tingkatan bahasa Jawa yang sesuai dalam komunikasi sehari-hari sehingga penggunaan Bahasa Jawa dapat dilestarikan dengan baik sesuai dengan norma kesantunan berbahasa.

### 3.5 Keterbatasan Penelitian

Penelitian ini mempunyai keterbatasan pada ketidakseimbangan dataset, di mana data Ngoko lebih dominan sehingga memengaruhi akurasi prediksi Krama Inggil. Sistem juga hanya mengenali kata yang sudah ada di dataset, sehingga tidak dapat memproses kata baru yang belum terdata. Selain itu, pendekatan menggunakan n-gram dan TF-IDF hanya mempertimbangkan pola kata tanpa memahami makna atau konteks sosial kalimat secara utuh, sehingga klasifikasi masih terbatas pada analisis kata per kata.

### 4. Kesimpulan

Penelitian ini berhasil mengimplementasikan sistem deteksi dan klasifikasi tingkatan Bahasa Jawa (Ngoko, Madya, dan Krama Inggil) menggunakan metode N-Gram dan algoritma Naive Bayes berbasis website interaktif. Sistem dapat memproses kalimat berbahasa Jawa, menganalisis kata per kata, dan menentukan tingkatan bahasa dengan akurasi yang cukup baik, terutama pada kelas Ngoko dan Madya dengan akurasi keseluruhan sebesar 83%. Sistem menunjukkan precision dan recall tinggi pada kelas Ngoko, cukup baik pada kelas Madya, namun masih rendah pada kelas Krama Inggil karena ketidakseimbangan data pada dataset yang digunakan. Sistem ini dapat membantu dalam pelestarian Bahasa Jawa serta menjadi media pembelajaran bagi masyarakat untuk memahami tingkatan Bahasa Jawa secara praktis. Meskipun

demikian, pengembangan lebih lanjut diperlukan untuk menyeimbangkan distribusi data dan meningkatkan akurasi prediksi pada kelas Krama Inggil agar sistem dapat digunakan secara optimal pada berbagai konteks komunikasi berbahasa Jawa.

Sebagai saran untuk penelitian selanjutnya, perlu dipertimbangkan penggunaan metode representasi teks yang lebih mampu menangkap konteks, seperti Word2Vec, FastText, atau BERT, sehingga sistem tidak hanya mengandalkan frekuensi kemunculan kata sebagaimana pada TF-IDF, tetapi juga dapat memahami hubungan semantik antar kata. Dengan pendekatan ini, model diharapkan dapat lebih peka terhadap perbedaan makna yang halus antar tingkatan bahasa Jawa. Selain itu, penambahan jumlah data pada kelas Krama Inggil serta penerapan teknik balancing data yang lebih optimal juga disarankan agar akurasi prediksi pada semua kelas dapat meningkat secara merata.

### Daftar Pustaka:

- Angeline, G., Wibawa, A. P., & Pujiyanto, U. (2022). Klasifikasi Dialek Bahasa Jawa Menggunakan Metode Naives Bayes. *Jurnal Mnemonic*, 5(2), 103–110. <https://doi.org/10.36040/mnemonic.v5i2.4748>
- Arasy, A., & Agustian, S. (2025). *Sentiment Classification Using Multilayer Perceptron Algorithm with TF-IDF Features Klasifikasi Sentimen Menggunakan Metode Multilayer Perceptron dengan Fitur TF-IDF*. 5(July), 908–919.
- Cahyo Prianto, Markuci, D., & Pane, S. F. (2023). Implementasi Spelling Corrector Untuk Mengatasi Typographical Error Pada Fitur Pencarian Aplikasi Kamus Istilah Informatika. *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, 17(1), 20–30. <https://doi.org/10.47111/jti.v17i1.5520>
- Dewi, P. N., Irawan, D., & Ulumuddin, I. (2025). *Perancangan Aplikasi Masdika Untuk Pembelajaran Bahasa Jawa Krama*. 4(1), 8–14.
- Garcia, A. R., Filipe, S. B., Fernandes, C., Estevão, C., & Ramos, G. (n.d.). *No 主觀的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title*. 9(April 2025), 1–6.
- Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40.
- Gita Segara, K., & Irwan Padli Nasution, M. (2025). Perkembangan Teknologi Informasi di Indonesia: Tantangan dan Peluang. *Jurnal Sains Student Research*, 3(1), 21–33. <https://doi.org/10.61722/jssr.v3i1.3128>
- Haris, A., Sandi, K., Haerani, E., Oktavia, L., &

- Kurnia, F. (2025). *BULLETIN OF COMPUTER SCIENCE RESEARCH Klasifikasi Sentimen Masyarakat Terhadap Revisi Undang-Undang Tentara Nasional Indonesia Menggunakan Naïve Bayes Classifier*. 5(4), 594–603. <https://doi.org/10.47065/bulletincsr.v5i4.615>
- Hidayatullah, A. F., Apong, R. A., Lai, D. T. C., & Qazi, A. (2023). Corpus creation and language identification for code-mixed Indonesian-Javanese-English Tweets. *PeerJ Computer Science*, 9, 1–24. <https://doi.org/10.7717/PEERJ-CS.1312>
- Nauli, S., Berutu, S., Budiati, H., & Maedjaja, F. (2025). JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Journal homepage: <https://jurnal.stkipgritulungagung.ac.id/index.php/jipi> KLASIFIKASI KALIMAT PERUNDUNGAN PADA TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(1), 107–122. <https://doi.org/10.29100/jipi.v10i1.5749>
- Nurhidayat, R., & Dewi, K. E. (2023). Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek. *Komputa : Jurnal Ilmiah Komputer dan Informatika*, 12(1), 91–100. <https://doi.org/10.34010/komputa.v12i1.9458>
- Puji Arfianingrum. (2020). Penerapan Unggah-Ungguh Bahasa Jawa Sesuai Dengan Konteks Tingkat Tutar Budaya Jawa. *Jurnal Prakarsa Paedagogia*, 03(02), 139–140.
- Putri, S. D. A., Ibrohim, M. O., & Budi, I. (2021). Abusive language and hate speech detection for Javanese and Sundanese languages in tweets: Dataset and preliminary study. *2021 11th International Workshop on Computer Science and Engineering, WCSE 2021, Wcse*, 461–465. <https://doi.org/10.18178/wcse.2021.02.011>
- Putri, U. S., Taryo, T., & Hindasyah, A. (2025). *Analisis Pemetaan Kemampuan Akademik Mahasiswa Program Studi Teknik Informatika Dengan Metode Naive Bayes Classifier Dan K-Nearest Neighbor ( KNN ) ( Studi Kasus : Universitas Pamulang )*. 7(1), 24–31.
- Rayuwati, Husna Gemasih, & Irma Nizar. (2022). IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID. *Jurnal Riset Rumpun Ilmu Teknik*, 1(1), 38–46. <https://doi.org/10.55606/jurritek.v1i1.127>
- Sentimen, A., Di, M., Sosial, M., Terhadap, X., Dengan, K., Bayes, N., Svm, D., Ryandi, F. A., Pratiwi, D., & Sari, S. (2025). Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM. *Jurnal Sains dan Teknologi*, 7(1), 1–6.
- SLAM, B. E., Feri Irawan, Nolan Efranda, & Rifaldi Herikson. (2025). Implementasi Machine Learning untuk Klasifikasi Buku Otomatis pada Perpustakaan Digital. *Jurnal Informatika Polinema*, 11(3), 305–310. <https://doi.org/10.33795/jip.v11i3.7298>
- Zaman, B., Hariyanti, E., Purwanti, E., & Bahasa, A. D. (2015). *Sistem Deteksi Bahasa pada Dokumen menggunakan N-Gram*. 1(2), 21–26.

*Halaman ini sengaja dikosongkan*

---