

---

## Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Berita Online Terhadap Program Prioritas Pemerintah Kota Medan

**M Choirul Amri<sup>1\*</sup>), Muhammad Farros Atiqi<sup>2)</sup> Friza Talita<sup>3)</sup> & Ilka Zufria<sup>4)</sup>**

<sup>1,2,3,4)</sup>Ilmu Komputer, Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara,

\*Coresponding Email: mchoirulamri112@gmail.com

---

### Abstrak

Penelitian ini bertujuan untuk mengimplementasikan algoritma Naïve Bayes dalam mengklasifikasikan berita online terkait Pemerintah Kota Medan ke dalam kategori Prioritas dan Non-Prioritas. Dataset yang digunakan terdiri dari 207 judul berita yang dihimpun dari berbagai sumber online. Setiap judul berita diproses melalui tahapan preprocessing seperti case folding, tokenizing, stopword removal, dan stemming untuk menghasilkan representasi fitur biner yang digunakan sebagai input klasifikasi. Pembagian dataset dilakukan dengan komposisi 70% data latih dan 30% data uji untuk menilai performa model secara objektif. Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa algoritma Naïve Bayes mampu mencapai akurasi sebesar 85,71% dalam mengelompokkan berita ke dalam kategori yang sesuai. Temuan ini menunjukkan bahwa metode Naïve Bayes cukup efektif dan dapat diimplementasikan sebagai alat bantu dalam mengidentifikasi berita yang berkaitan dengan program-program prioritas Pemerintah Kota Medan secara otomatis.

**Kata Kunci:** Naïve bayes, Klasifikasi berita, Program Prioritas Pemerintah.

### Abstract

*This study aims to implement the Naïve Bayes algorithm to classify online news related to the Medan City Government into Priority and Non-Priority categories. The dataset consists of 207 news headlines collected from various online sources. Each headline undergoes several preprocessing stages, including case folding, tokenizing, stopword removal, and stemming, to produce binary feature representations used for classification. The dataset is divided into 70% training data and 30% testing data to objectively evaluate model performance. Based on the confusion matrix, the Naïve Bayes algorithm achieves an accuracy of 85.71% in categorizing news headlines into their appropriate classes. These findings indicate that the Naïve Bayes method is effective and can be implemented as a supporting tool to automatically identify news related to the priority programs of the Medan City Government.*

**Keywords:** Naïve Bayes, News Classification, Government Priority Programs.

---

## PENDAHULUAN

Kehadiran internet mengubah cara manusia saling berinteraksi. Pesatnya perkembangan teknologi komunikasi pun mendorong perubahan hubungan antara teknologi, industri, dan konsumsi media (Suciska et al., 2020). Saat ini, informasi lebih banyak diperoleh melalui media daring dan media sosial dibandingkan media konvensional seperti televisi atau surat kabar. Di Indonesia, konsumsi berita online dan melalui media sosial menjadi yang tertinggi dibandingkan kanal lainnya (Anugrafianto & Artikel, 2023). Hal ini menandakan adanya pergeseran perilaku masyarakat dalam memperoleh informasi, sekaligus memperlihatkan potensi peredaran informasi yang masif, cepat, dan beragam kualitasnya (Kurnia Ilahi et al., 2021).

Perubahan tersebut menuntut instansi pemerintah untuk mampu beradaptasi dalam mengelola dan memantau informasi publik secara lebih efisien. Dinas Komunikasi dan Informatika (Diskominfo) Kota Medan sebagai instansi yang berperan penting dalam penyebaran informasi publik memiliki tanggung jawab besar untuk memastikan bahwa data dan kegiatan publikasi pemerintah daerah dapat dikelola serta dimonitor secara akurat dan transparan. Kegiatan utama pada Divisi Komunikasi Publik meliputi pengelolaan kliping berita dari berbagai media massa yang memuat aktivitas Wali Kota Medan, publikasi program pembangunan, serta kegiatan prioritas daerah.

Namun, dalam pelaksanaannya, proses penginputan berita kegiatan Wali Kota Medan saat ini masih dilakukan melalui Google Form. Setiap berita yang masuk kemudian dikategorikan secara manual ke dalam label program prioritas atau non-prioritas. Proses manual ini membutuhkan waktu yang cukup lama dan rentan terhadap ketidakkonsistenan, terutama karena jumlah data berita yang diinput setiap hari sangat besar dan terus bertambah. Selain itu, variasi bahasa dan topik berita membuat proses klasifikasi manual menjadi semakin kompleks dan berpotensi menimbulkan kesalahan kategori. Kondisi tersebut menghambat efektivitas proses pemantauan serta memperlambat penyusunan laporan evaluasi publikasi media yang berkaitan dengan kinerja Pemerintah Kota Medan.

Untuk mengatasi permasalahan tersebut, diperlukan sistem dengan pendekatan berbasis machine learning yang mampu melakukan klasifikasi berita secara otomatis dan akurat. Salah satu algoritma yang banyak digunakan untuk tugas klasifikasi teks adalah Naive Bayes, yang dikenal karena kesederhanaan, efisiensi, dan performanya yang baik dalam pengolahan data teks (text mining)(Irsyad Ramadhan & Zufria, 2024). Algoritma ini bekerja berdasarkan teorema Bayes dengan asumsi independensi antar fitur, sehingga mampu memberikan hasil yang optimal dalam waktu pemrosesan yang relatif singkat (Purnama & Wahyudi, 2024).

Sejumlah penelitian terdahulu telah menunjukkan efektivitas algoritma Naive Bayes dalam konteks klasifikasi teks dan berita. (Febriyanty et al., 2023) membandingkan kinerja Naive Bayes dan Support Vector Machine dalam deteksi berita hoaks berbahasa Indonesia dan menemukan bahwa Naive Bayes memiliki akurasi mencapai 88%, lebih tinggi dibandingkan SVM (75,5%). Penelitian lain oleh (Bhakti et al., 2024) juga berhasil mengimplementasikan Naive Bayes untuk klasifikasi kategori berita umum dengan hasil akurasi yang stabil. (Tarigan et al., 2024) membuktikan bahwa dengan penerapan pembobotan TF-IDF, Naive Bayes dapat mencapai akurasi hingga 97% dalam mendeteksi berita hoaks. Selain itu, (Pramudita et al., 2024) menunjukkan bahwa Naive Bayes dapat digunakan untuk analisis sentimen publik terhadap program pemerintah, dengan akurasi mencapai 84,99%.

Dari berbagai penelitian tersebut dapat disimpulkan bahwa algoritma Naive Bayes memiliki performa yang kuat dalam menangani klasifikasi teks berbahasa Indonesia, baik dalam konteks deteksi hoaks, analisis sentimen, maupun kategorisasi berita. Namun, sebagian besar penelitian sebelumnya masih berfokus pada klasifikasi hoaks atau sentimen publik, bukan pada klasifikasi berita kebijakan daerah berdasarkan program prioritas pemerintah. Oleh karena itu, terdapat kesenjangan riset yang dapat diisi melalui penerapan *Naive Bayes* dalam konteks klasifikasi berita program prioritas Pemerintah Kota Medan.

Penelitian ini bertujuan untuk mengimplementasikan algoritma Naive Bayes dalam klasifikasi berita online yang berkaitan dengan kegiatan Wali Kota Medan berdasarkan sepuluh program prioritas Pemerintah Kota Medan. Dengan adanya sistem ini, proses penginputan dan pengelompokan berita yang semula dilakukan secara manual dapat digantikan dengan sistem klasifikasi otomatis yang lebih cepat, akurat, dan konsisten. Secara teoretis, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan penerapan machine learning dalam analisis kebijakan publik berbasis teks. Sedangkan secara praktis, hasil penelitian ini dapat menjadi dasar pengembangan sistem pemantauan media berbasis web yang membantu Diskominfo Kota Medan dalam mengelola data publikasi secara efektif, meningkatkan efisiensi kerja, serta memperkuat akuntabilitas komunikasi publik pemerintah daerah.

## **METODE PENELITIAN**

### **A. Pendekatan Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode machine learning supervised classification. Fokus utama penelitian adalah mengimplementasikan dan menguji performa algoritma Naive Bayes dalam

melakukan klasifikasi teks berita daring ke dalam kategori Prioritas dan Non Prioritas.

Pendekatan ini dipilih karena mampu menghasilkan model klasifikasi berbasis probabilistik yang dapat memprediksi kategori berita baru berdasarkan pola kata dari data pelatihan (*training data*) (Wibowo et al., 2024). Proses eksperimen dilakukan melalui tahapan pengumpulan data, praproses teks, pelabelan data, pelatihan model, dan evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-score (Soleman, 2021).

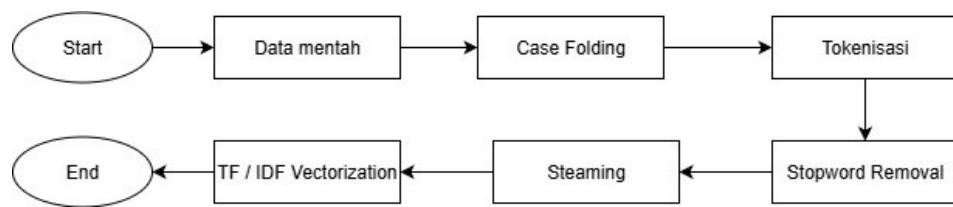
## B. Sumber Data

Data yang digunakan dalam penelitian ini berupa judul dan isi berita daring yang diperoleh dari beberapa portal berita lokal maupun nasional yang secara rutin meliput kegiatan Pemerintah Kota Medan. Portal-portal berita tersebut meliputi media seperti Medan Pos, Analisa Daily, Tribun Medan, Waspada Online, dan beberapa situs berita lain yang menayangkan publikasi resmi terkait aktivitas Wali Kota Medan dan program-program prioritas Pemerintah Kota Medan.

Proses pengumpulan data dilakukan secara manual, yaitu dengan mencari, menyeleksi, dan menyalin berita-berita yang relevan dari berbagai sumber daring (Hartono et al., 2023). Berita yang dikumpulkan merupakan publikasi yang memuat kegiatan Pemerintah Kota Medan, khususnya yang berkaitan dengan sepuluh program prioritas Wali Kota Medan. Dari hasil proses pengumpulan tersebut, diperoleh sebanyak 207 berita daring yang kemudian digunakan sebagai dataset dalam penelitian ini.

## C. Proses Praproses Teks

Proses praproses dalam penelitian ini mencakup beberapa langkah utama, yaitu case folding, tokenisasi, stopword removal, stemming, dan TF-IDF vectorization (Setiawan et al., 2025). Langkah-langkah ini dilakukan secara berurutan agar menghasilkan data yang terstruktur dengan baik dan siap digunakan pada tahap pelatihan model Naive Bayes (Rahutomo et al., 2019). Proses tersebut divisualisasikan pada Gambar 1, yang memperlihatkan alur mulai dari data berita mentah hingga terbentuk representasi vektor numerik hasil TF-IDF yang akan menjadi input bagi model klasifikasi (Denny et al., 2019).



Gambar 1 Pre-processing Stage

#### D. Labeling dan Kategori Kelas

Proses pelabelan data dilakukan berdasarkan kesesuaian isi berita dengan sepuluh program prioritas Pemerintah Kota Medan, seperti revitalisasi pasar tradisional, pengembangan ruang terbuka hijau, transportasi publik terintegrasi, pengelolaan sampah, pembangunan drainase, pusat kreativitas anak muda, penanggulangan stunting, pengembangan pariwisata, penyediaan air bersih dan sanitasi, serta digitalisasi pendidikan. Berita yang memiliki keterkaitan dengan salah satu program tersebut diberi label "Prioritas", sedangkan berita lain yang tidak relevan secara langsung diberi label "Non-Prioritas". Pelabelan dilakukan secara manual oleh dua validator independen dengan meninjau isi berita untuk memastikan kesesuaian konteks. Hasil pelabelan kemudian dibandingkan menggunakan inter-rater agreement guna mengukur tingkat konsistensi antar penilai, dan dataset akhir ini digunakan dalam proses pelatihan serta pengujian model Naive Bayes (Febriyanty et al., 2023).

#### E. Naïve Bayes Classification

Naive Bayes Classifier adalah suatu metode klasifikasi yang berasal dari teorema Bayes. Metode ini menggunakan probabilitas dan statistik untuk melakukan klasifikasi. Naive Bayes Classifier memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Rieuwpassa et al., 2024). Model bekerja berdasarkan rumus teorema Bayes berikut:

di mana:

- $P(C|X)$  adalah probabilitas kelas C diberikan fitur X,
- $P(X|C)$  adalah probabilitas munculnya fitur pada kelas tersebut,
- $P(C)$  adalah probabilitas awal kelas, dan
- $P(X)$  adalah probabilitas fitur secara keseluruhan.

Model ini mengasumsikan bahwa setiap kata pada dokumen bersifat independen terhadap kata lainnya. Meskipun asumsi tersebut sederhana, berbagai studi membuktikan efektivitasnya untuk tugas klasifikasi berita dan analisis sentimen (Cahyani & Budiman, 2025).

### HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Pada tahap ini dijelaskan sumber dan jenis data yang digunakan dalam penelitian. Data yang dikumpulkan berupa judul-judul berita kegiatan Pemerintah Kota Medan yang diambil dari portal berita daring lokal. Setiap berita diberi label prioritas atau non-prioritas berdasarkan kesesuaiannya dengan sepuluh program

prioritas Pemerintah Kota Medan, seperti revitalisasi pasar tradisional, pengembangan RTH, transportasi publik, dan sebagainya. Dataset disimpan dalam format Excel dan dibagi menjadi data latih (70%) dan data uji (30%).

Tabel 1 Data Mentah

No	Judul berita
1	Warga Mabar Hilir Curhat Soal Banjir Lingkungan, Rico Waas: Kita Segera Tuntaskan
2	Taman Hening dan Gajah Mada Dibenahi, Wali Kota Medan Rico Waas Berencana Tambah Jogging Track
3	Nobar Timnas vs Vietnam di Balai Kota Medan, Rico Waas: Pertahanan Vietnam Ketat Seperti Kanebo Baru
4	Hadiri Maulid Nabi, Rico Waas Tegaskan Ketenteraman Keluarga Kunci Kedamaian Medan
.....	.....
206	Rico Waas Targetkan Taman Kota yang Ramah Bagi Anak
207	Wali Kota Medan Rico Waas Tutup Porkot XV 2025, Hadirkan Atraksi dan Lucky Draw Berhadiah Motor

## B. Tahap Pre-processing

Sebelum digunakan untuk klasifikasi, dataset terlebih dahulu diolah melalui tahap pre-processing. Tahap ini merupakan bagian penting dari proses data mining yang bertujuan menyiapkan data mentah agar layak untuk dianalisis. Dalam penelitian ini, data berupa judul-berita dari portal daring Kota Medan dibersihkan dari kesalahan agar analisis klasifikasi terhadap program prioritas Pemerintah Kota Medan memberikan hasil yang akurat dan relevan.

### 1. Case Folding

Tahap pertama akan dilakukan Transform Cases, proses ini bertujuan untuk menyeragamkan penulisan huruf pada seluruh data teks agar tidak terjadi perbedaan makna antara kata yang sebenarnya sama tetapi ditulis dengan bentuk huruf berbeda (Cindy Astuti et al., 2024), hasil dari proses ini ditunjukkan pada table 2

Tabel 2 Case folding

No	sebelum	sesudah
1	Warga Mabar Hilir Curhat Soal Banjir Lingkungan, Rico Waas: Kita Segera Tuntaskan	warga mabar hilir curhat soal banjir lingkungan rico waas kita segera tuntaskan

2	Taman Hening dan Gajah Mada Dibenahi, Wali Kota Medan Rico Waas Berencana Tambah Jogging Track	taman hening dan gajah mada dibenahi wali kota medan rico waas berencana tambah jogging track
3	Nobar Timnas vs Vietnam di Balai Kota Medan, Rico Waas: Pertahanan Vietnam Ketat Seperti Kanebo Baru	nobar timnas vs vietnam di balai kota medan rico waas pertahanan vietnam ketat seperti kanebo baru
4	Hadiri Maulid Nabi, Rico Waas Tegaskan Ketenteraman Keluarga Kunci Kedamaian Medan	hadiri maulid nabi rico waas tegaskan ketenteraman keluarga kunci kedamaian medan

## 2. Tokenisasi

Proses berikutnya adalah tokenizing atau tokenisasi,. Tujuan dari tokenisasi adalah agar sistem dapat memahami teks bukan sebagai satu kalimat utuh, melainkan sebagai kumpulan kata yang dapat diolah secara terpisah (Yoga Pratama et al., 2025), hasil dari proses ini ditunjukkan pada table 3.

Tabel 3 Tokenisasi

No	sebelum	sesudah
1	warga mabar hilir curhat soal banjir lingkungan rico waas kita segera tuntaskan	'warga', 'mabar', 'hilir', 'curhat', 'soal', 'banjir', 'lingkungan', 'rico', 'waas', 'kita', 'segera', 'tuntaskan'
2	taman hening dan gajah mada dibenahi wali kota medan rico waas berencana tambah jogging track	'taman', 'hening', 'dan', 'gajah', 'mada', 'dibenahi', 'wali', 'kota', 'medan', 'rico', 'waas', 'berencana', 'tambah', 'jogging', 'track'
3	nobar timnas vs vietnam di balai kota medan rico waas pertahanan vietnam ketat seperti kanebo baru	'nobar', 'timnas', 'vs', 'vietnam', 'di', 'balai', 'kota', 'medan', 'rico', 'waas', 'pertahanan', 'vietnam', 'ketat', 'seperti', 'kanebo', 'baru'
4	hadiri maulid nabi rico waas tegaskan ketenteraman keluarga kunci kedamaian medan	'hadiri', 'maulid', 'nabi', 'rico', 'waas', 'tegaskan', 'ketenteraman', 'keluarga', 'kunci', 'kedamaian', 'medan'

## 3. Stopword Removal

Tahap selanjutnya adalah stopwords removal, stopword removal dilakukan untuk membuang kata dalam data teks yang sering muncul dan tidak berhubungan dengan teks (Zufria et al., 2024). hasil dari proses ini ditunjukkan pada table 4

Tabel 4 Stopword Removal

No	sebelum	sesudah
1	'warga', 'mabar', 'hilir', 'curhat', 'soal', 'banjir', 'lingkungan', 'rico', 'waas', 'kita', 'segera', 'tuntaskan'	'warga' 'mabar' 'hilir' 'curhat' 'banjir' 'lingkung' 'rico' 'waas' 'tuntaskan'
2	'taman', 'hening', 'dan', 'gajah', 'mada', 'dibenahi', 'wali', 'kota', 'medan', 'rico', 'waas', 'berencana', 'tambah', 'jogging', 'track'	'taman' 'hening' 'gajah' 'mada' 'benahi' 'wali' 'kota' 'medan' 'rico' 'waas' 'rencana' 'tambah' 'jogging' 'track'
3	'nobar', 'timnas', 'vs', 'vietnam', 'di', 'balai', 'kota', 'medan', 'rico', 'waas', 'pertahanan', 'vietnam', 'ketat', 'seperti', 'kanebo', 'baru'	'nobar' 'timnas' 'vietnam' 'balai' 'kota' 'medan' 'rico' 'waas' 'tahan' 'vietnam' 'ketat' 'kanebo' 'baru'
4	'hadiri', 'maulid', 'nabi', 'rico', 'waas', 'tegaskan', 'ketenteraman', 'keluarga', 'kunci', 'kedamaian', 'medan'	'hadiri' 'maulid' 'nabi' 'rico' 'waas' 'tegas' 'tentram' 'keluarga' 'kunci' 'damai' 'medan'

#### 4. Steamming

Selanjutnya adalah proses stemming, yaitu tahap untuk mengubah setiap kata ke bentuk dasarnya (root word). Tujuan utama dari stemming adalah menyederhanakan variasi bentuk kata agar sistem mengenal kata dengan makna yang sama sebagai satu entitas tunggal, hasil dari proses ini ditunjukkan pada table 5

Tabel 5 Steamming

No	sebelum	sesudah
1	'warga' 'mabar' 'hilir' 'curhat' 'banjir' 'lingkungan' 'rico' 'waas' 'tuntaskan'	'warga' 'mabar' 'hilir' 'curhat' 'banjir' 'lingkung' 'rico' 'waas' 'tuntas'
2	'taman' 'hening' 'gajah' 'mada' 'benahi' 'wali' 'kota' 'medan' 'rico' 'waas' 'rencana' 'tambah' 'jogging' 'track'	'taman' 'hening' 'gajah' 'mada' 'benah' 'wali' 'kota' 'medan' 'rico' 'waas' 'rencana' 'tambah' 'jogging' 'track'
3	'nobar' 'timnas' 'vietnam' 'balai' 'kota' 'medan' 'rico' 'waas' 'tahan' 'vietnam' 'ketat' 'kanebo' 'baru'	'nobar' 'timnas' 'vietnam' 'balai' 'kota' 'medan' 'rico' 'waas' 'tahan' 'vietnam' 'ketat' 'kanebo' 'baru'
4	'hadiri' 'maulid' 'nabi' 'rico' 'waas' 'tegas' 'tentram' 'keluarga' 'kunci' 'damai' 'medan'	'hadir' 'maulid' 'nabi' 'rico' 'waas' 'tegas' 'tentram' 'keluarga' 'kunci' 'damai' 'medan'

## 5. Pembobotan TF-IDF

Dari dokumen pertama, kata “banjir” muncul sebanyak 1 kali dari total 12 kata. Maka perhitungan nilai Term Frequency (TF) adalah sebagai berikut:

$$TF_{\text{Banjir}} = \frac{1}{12} = 0.0833$$

Selanjutnya dilakukan perhitungan nilai Inverse Document Frequency (IDF). Berikut adalah langkah perhitungannya:

$$IDF_{\text{banjir}} = \log \left( \frac{1}{4} \right) \log (4) = 0.6020$$

Setelah memperoleh nilai TF dan IDF, langkah berikutnya adalah menghitung nilai TF-IDF dengan mengalikan kedua nilai tersebut:

$$TFIDF_{\text{banjir}} = 0.0833 \times 0.6020 = 0.0501$$

Perhitungan di atas disajikan sebagai contoh untuk menggambarkan proses pembobotan TF-IDF pada salah satu kata dalam satu dokumen. Nilai TF, IDF, dan TF-IDF yang diperoleh hanya bertujuan memberikan ilustrasi mengenai tahapan perhitungannya, sementara perhitungan sebenarnya pada penelitian ini dilakukan pada seluruh kata dalam keseluruhan dataset agar menghasilkan bobot yang lebih komprehensif dan akurat.

## C. Pengujian dan Hasil

### 1. Implementasi Algoritma Naïve Bayes

Dalam penelitian ini, algoritma Naïve Bayes digunakan untuk mengklasifikasikan 207 data latih berupa judul berita ke dalam kategori Prioritas dan Non-Prioritas berdasarkan distribusi probabilitas dari fitur biner. Setiap judul direpresentasikan dengan nilai 1 atau 0 sesuai kemunculan kata tertentu. Untuk contoh perhitungan, dipilih 4 data latih secara acak agar variasi kata dalam dataset tetap terwakili. Data latih tersebut ditampilkan pada table 6 berikut.

Tabel 6 Data uji

No	Judul	Clean Text	Label
1	Warga Mabar Hilir Curhat Soal Banjir Lingkungan, Rico Waas: Kita Segera Tuntaskan	‘warga’ ‘mabar’ ‘hilir’ ‘curhat’ ‘soal’ ‘banjir’ ‘lingkung’ ‘rico’ ‘waas’ ‘kita’ ‘segera’ ‘tuntas’	Prioritas
2	Taman Hening dan Gajah Mada Dibenahi, Wali Kota Medan Rico Waas Berencana Tambah Jogging Track	‘taman’ ‘hening’ ‘dan’ ‘gajah’ ‘mada’ ‘benah’ ‘wali’ ‘kota’ ‘medan’ ‘rico’ ‘waas’ ‘rencana’ ‘tambah’ ‘jogging’ ‘track’	Prioritas
3	Nobar Timnas vs Vietnam di Balai Kota Medan, Rico	‘nobar’ ‘timnas’ ‘vs’ ‘vietnam’ ‘di’ ‘balai’ ‘kota’ ‘medan’ ‘rico’ ‘waas’	Non Prioritas

	Waas: Pertahanan Vietnam Ketat Seperti Kanebo Baru	'tahan' 'vietnam' 'ketat' 'seperti' 'kanebo' 'baru'	
4	Hadiri Maulid Nabi, Rico Waas Tegaskan Ketenteraman Keluarga Kunci Kedamaian Medan	'hadir' 'maulid' 'nabi' 'rico' 'waas' 'tegas' 'tentram' 'keluarga' 'kunci' 'damai' 'medan'	Non Prioritas
Data Uji	Rico Waas Merenofasi Taman Cadika	'rico' 'waas' 'renovasi' 'taman' 'cadika'	?

Pada tahapan klasifikasi, proses penentuan kelas diawali dengan menghitung prior probability untuk masing-masing kategori.

$$P(\text{Kelas kategori}) = \frac{\text{Jumlah Kelas X}}{\text{Jumlah Kategori}}$$

Dengan menggunakan persamaan diatas, maka probabilitas dari setiap kelas pada kategori adalah

$$a. P(\text{Prioritas|Kategori}) = \frac{2}{4} = 0,5$$

$$b. P(\text{Non Prioritas|Kategori}) = \frac{2}{4} = 0,5$$

Perhitungan nilai conditional probability

Dengan persamaan rumus di atas maka dapat dihitung probabilitas dari term pada setiap kelas pada kategori.

a. Prioritas

$$P(\text{rico|Prioritas}) = \frac{2+1}{2+2} = \frac{3}{4} = 0,75$$

$$P(\text{waas|Prioritas}) = \frac{2+1}{2+2} = \frac{3}{4} = 0,75$$

$$P(\text{renovasi|Prioritas}) = \frac{0+1}{2+2} = \frac{3}{4} = 0,25$$

$$P(\text{taman|Prioritas}) = \frac{1+1}{2+2} = \frac{3}{4} = 0,50$$

$$P(\text{cadika|Prioritas}) = \frac{0+1}{2+2} = \frac{3}{4} = 0,25$$

b. Non Prioritas

$$P(rico|Non\ Prioritas) = \frac{2+1}{2+2} = \frac{3}{4} = 0,75$$

$$P(waas|Non\ Prioritas) = \frac{2+1}{2+2} = \frac{3}{4} = 0,75$$

$$P(renovasi|Non\ Prioritas) = \frac{0+1}{2+2} = \frac{3}{4} = 0,25$$

$$P(taman|Non\ Prioritas) = \frac{0+1}{2+2} = \frac{3}{4} = 0,25$$

$$P(cadika|Non\ Prioritas) = \frac{0+1}{2+2} = \frac{3}{4} = 0,25$$

Posterior Prioritas

$$P(Prioritas|X) = P(Prioritas) \times (0.75 \times 0.75 \times 0.25 \times 0.50 \times 0.25)$$

$$= 0.5 \times 0.017578125$$

$$= 0.0087890625$$

Posterior Non Prioritas

$$P(Non\ Prioritas|X) = P(Non\ Prioritas) \times (0.75 \times 0.75 \times 0.25 \times 0.25 \times 0.25)$$

$$= 0.5 \times 0.0087890625$$

$$= 0.00439453125$$

Berdasarkan hasil perhitungan menggunakan algoritma Bernoulli Naïve Bayes, berita uji "Rico Waas Merenovasi Taman Cadika" diklasifikasikan ke dalam kategori Prioritas.

## 2. Pengujian Model

Pada tahap pengujian model, dari total 207 data, dilakukan pembagian 70% data latih dan 30% data uji, sehingga diperoleh 42 data uji untuk mengevaluasi kinerja klasifikasi. Pembagian ini memastikan bahwa model tidak hanya belajar dari data latih, tetapi juga diuji menggunakan data baru sehingga hasil prediksinya dapat dinilai secara objektif untuk menghitung nilai akurasi dapat ditentukan dengan rumus dibawah ini.

$$Akurasi = \frac{T\ prioritas + T\ non\ prioritas}{total\ data\ uji}$$

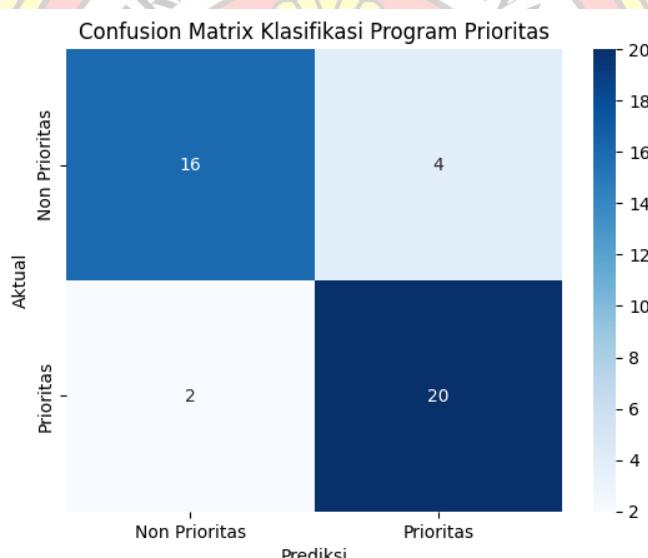
Setelah dilakukan pengujian model pada data uji maka hasil yang di peroleh adalah sebagai berikut:

• Akurasi: 0.8571428571428571
Laporan Klasifikasi:
precision recall f1-score support
Non Prioritas 0.89 0.80 0.84 20
Prioritas 0.83 0.91 0.87 22
accuracy 0.86 0.85 0.86 42
macro avg 0.86 0.85 0.86 42
weighted avg 0.86 0.86 0.86 42

Gambar 2 Akurasi data uji

### 3. Evaluasi Model

Pada tahap evaluasi model, hasil prediksi dibandingkan dengan label sebenarnya menggunakan confusion matrix untuk melihat sejauh mana model dapat mengklasifikasikan data dengan benar. Matriks ini menampilkan jumlah prediksi benar dan salah pada kategori Prioritas dan Non-Prioritas, sehingga memudahkan penilaian akurasi dan kinerja keseluruhan model.

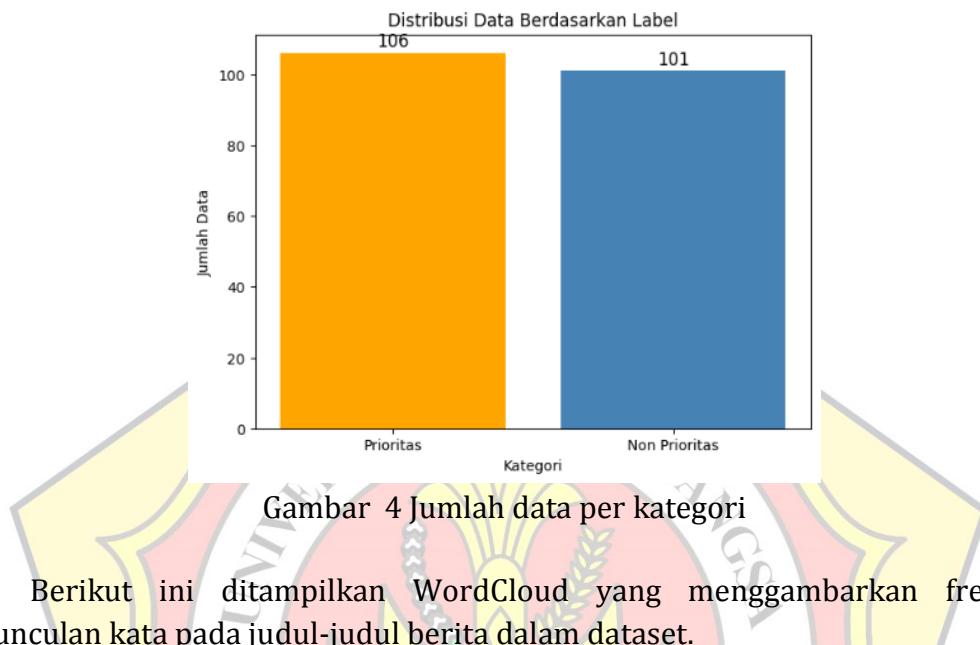


Gambar 3 Confussion Matriks pada data uji

Berdasarkan confusion matrix di atas, dapat dilihat bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, yaitu 16 data Non-Prioritas dan 20 data Prioritas yang terprediksi sesuai label sebenarnya. Sementara itu, masih terdapat beberapa kesalahan prediksi, yaitu 4 data Non-Prioritas yang salah diklasifikasikan sebagai Prioritas dan 2 data Prioritas yang diprediksi sebagai Non-Prioritas. Untuk mengetahui kinerja model secara keseluruhan, selanjutnya dilakukan perhitungan akurasi menggunakan rumus berikut.

$$Akurasi = \frac{16 + 20}{16 + 4 + 2 + 20} \times 100\% = 85,71\%$$

Berdasarkan hasil perhitungan manual menggunakan confusion matrix, diperoleh akurasi sebesar 85,71%. Nilai ini menunjukkan bahwa model berhasil mengklasifikasikan sebagian besar data uji dengan benar, sehingga dapat dikatakan bahwa kinerja model dalam membedakan berita Prioritas dan Non-Prioritas berada pada kategori baik.



Berikut ini ditampilkan WordCloud yang menggambarkan frekuensi kemunculan kata pada judul-judul berita dalam dataset.

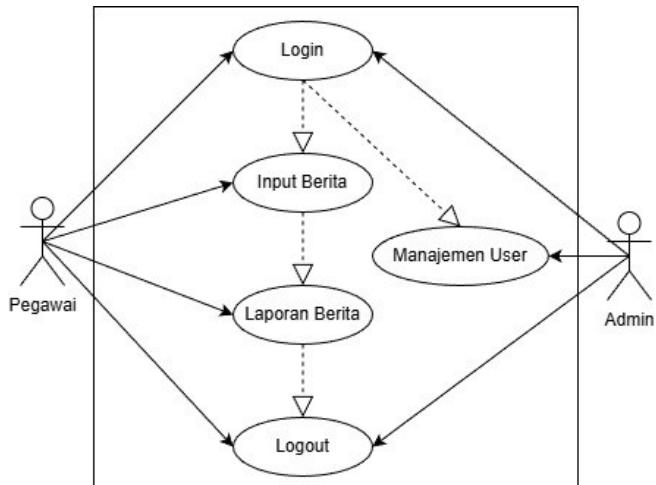


Gambar 5 Wordcloud judul berita

#### D. Rancangan ke Sistem

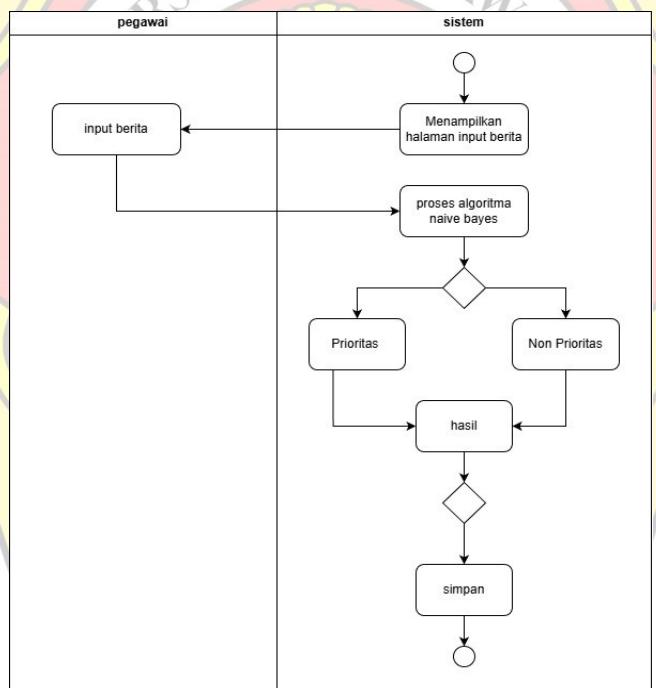
Berdasarkan analisis yang telah dilakukan terhadap metode Naïve Bayes dalam pemilihan kategori pada judul berita, perancangan sistem ini disusun untuk memastikan setiap tahapan klasifikasi berjalan secara terstruktur dan efisien, maka hasil dari perancangan sebagai berikut:

## 1. Use case diagaram



Gambar 6 use case diagram

## 2. Activity diagram



Gambar 7 Diagram activity

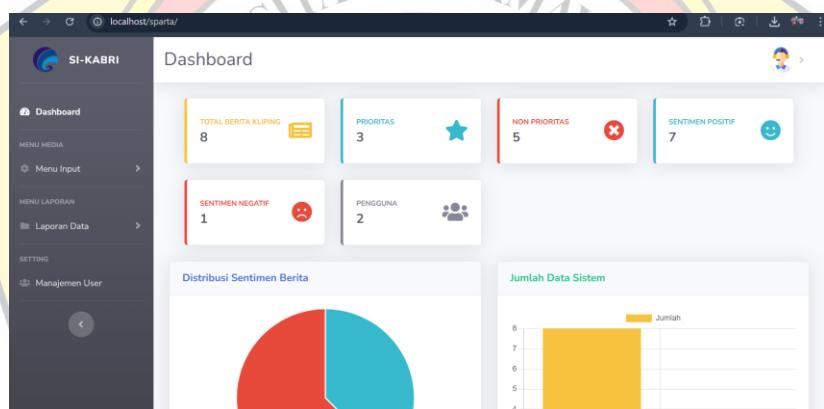
## 3. Desain Interface

Halaman login merupakan bagian awal yang harus diakses oleh pengguna sebelum memasuki sistem. Halaman ini berfungsi sebagai proses autentikasi untuk memastikan bahwa hanya pengguna yang memiliki hak akses yang dapat masuk ke dalam aplikasi.



Gambar 8 Halaman login

Setelah berhasil melakukan login, pengguna akan diarahkan ke halaman dashboard yang berfungsi sebagai pusat informasi utama di dalam sistem, tempat berbagai menu, ringkasan data, dan fitur penting ditampilkan secara ringkas dan mudah diakses.

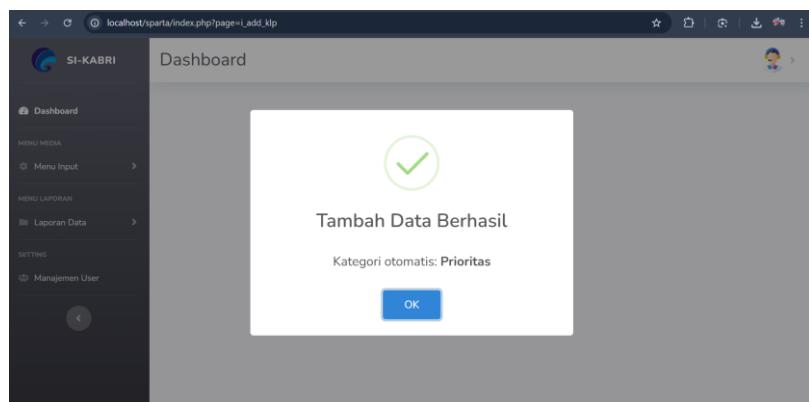


Gambar 9 Halaman dashboard

Pada halaman ini, pengguna dapat memasukkan informasi seperti tanggal, nama media, judul berita, serta menentukan sentimen berita.

Gambar 10 Halaman input berita

Sistem juga memberikan notifikasi bahwa klasifikasi kategori Prioritas atau Non Prioritas akan dilakukan secara otomatis.



Gambar 11 Notifikasi hasil

Data ini dapat dikelola oleh pengguna melalui fitur tambah, edit, dan hapus pada menu aksi, sehingga memudahkan proses verifikasi dan pengelolaan arsip berita secara efektif serta mendukung proses klasifikasi program prioritas Pemerintah Kota Medan.

No	Tanggal	Media	Judul Berita	Sentimen	Kategori	Aksi
1	19/Nov/2025	agiodeli.id	Wali kota medan merenovasi taman cadika	Positif	Prioritas	 
2	19/Nov/2025	agiodeli.id	wali kota medan hadiri Maulid nabi	Positif	Non Prioritas	 
3	16/Nov/2025	agiodeli.id	Rico Waas Saksikan Penyembelihan Hewan Kurban di	Positif	Non Prioritas	 

Gambar 12 Data yang tersimpan pada sistem

## SIMPULAN

Berdasarkan proses pengujian yang dilakukan, implementasi algoritma Naïve Bayes dalam mengklasifikasikan berita online terkait Program Prioritas Pemerintah Kota Medan menunjukkan kinerja yang cukup baik. Model yang dibangun mampu mengenali pola-pola teks pada berita sehingga dapat membedakan antara kategori Prioritas dan Non Prioritas secara otomatis. Dari hasil evaluasi menggunakan dataset yang telah melalui tahapan pra-pemrosesan, diperoleh nilai akurasi sebesar 85,71%, yang mengindikasikan bahwa algoritma Naïve Bayes efektif digunakan dalam sistem klasifikasi berita pada penelitian ini.

## DAFTAR PUSTAKA

- Anugrafianto, T. R., & Artikel, R. (2023). Analisis Dampak Media Digital terhadap Pola Konsumsi Berita Generasi Milenial di Indonesia. *Jurnal Ilmu Sosial Dan Humaniora*, 3(3), 21–25.
- Bhakti, I. N., Sholikhin, A. Z., Lukman, M. A., Daniati, E., & Ristyawan, A. (2024). Klasifikasi Kategori Berita Menggunakan Naive Bayes. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8, 1155–1163.
- Cahyani, O. N., & Budiman, F. (2025). Performa Logistic Regression dan Naive Bayes dalam Klasifikasi Berita Hoax di Indonesia. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 60–68. <https://doi.org/10.29408/edumatic.v9i1.28987>
- Cindy Astuti, K., Firmansyah, A., Riyadi, A., & Pelita Bangsa Bekasi, U. (2024). Implementasi Text Mining untuk Analisis Sentimen Masyarakat terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 8(1). <https://doi.org/10.33395/remik.v8i1.13421>
- Denny, Chandra, N., Indrawan, G., & Nyoman Sukajaya. (2019). KLASIFIKASI BERITA LOKAL RADAR MALANG MENGGUNAKAN METODE NAÏVE BAYES DENGAN FITUR N-GRAM. *Jurnal Ilmu Komputer Indonesia (JIKI)*, 4(2), 10–20.
- Febriyanty, N. E., Hariyadi, M. A., & Crysdian, C. (2023). Hoax Detection News Using Naïve Bayes and Support Vector Machine Algorithm. *International Journal of Advances in Data and Information Systems*, 4(2), 191–200. <https://doi.org/10.25008/ijadis.v4i2.1306>
- Hartono, H., Hajjah, A., & Marlism, Y. N. (2023). PENERAPAN METODE NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI JUDUL BERITA APPLICATION OF THE NAÏVE BAYES CLASSIFIER METHOD FOR NEWS TITLE CLASSIFICATION. *Jurnal SimanteC*, 12(1), 37–46.
- Irsyad Ramadhan, M., & Zufria, I. (2024). KLASIFIKASI KEAHlian INDIVIDU PEMAIN ONLINE GAMES DENGAN MENGGUNAKAN NAÏVE BAYES CLASSIFIER. *Journal of Science and Social Research*, 7(4), 2114–2120. <http://jurnal.goretanpena.com/index.php/JSSR>
- Kurnia Ilahi, F., Syafuddin, K., & Kunci, K. (2021). Konsumsi Pemberitaan Media Online di Masa Pandemi oleh Ikatan Pelajar Riau Yogyakarta (IPRY-KB) Komisariat Bengkalis Selama Covid-19. *Jurnal Riset Komunikasi*, 4, 275–292.
- Pramudita, D., Akbar, Y., & Wahyudi, T. (2024). Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah pada Media Sosial X Menggunakan Algoritma Naive Bayes. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1420–1430. <https://doi.org/10.57152/malcom.v4i4.1565>
- Purnama, L., & Wahyudi, T. (2024). Analisa Sentimen Tentang Piala Dunia u-20 Indonesia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Sains Dan Teknologi*, 6(2), 217–222. <https://doi.org/10.55338/saintek.v6i2.1397>
- Rahutomo, F., Retno, A., & Ririd, T. H. (2019). EVALUASI DAFTAR STOPWORD BAHASA INDONESIA. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 6(1), 41–48. <https://doi.org/10.25126/jtiik.201861226>
- Rieuwpassa, J. A., Sugito, S., & Widiharih, T. (2024). IMPLEMENTASI METODE NAIVE BAYES CLASSIFIER UNTUK KLASIFIKASI SENTIMEN ULASAN PENGGUNA

- APLIKASI NETFLIX PADA GOOGLE PLAY. *Jurnal Gaussian*, 12(3), 362–371. <https://doi.org/10.14710/j.gauss.12.3.362-371>
- Setiawan, A., Abidin, Z., & Imamudin, M. (2025). Impact of Preprocessing on Indonesian Extractive Summarization Using LexRank, TextRank, DivRank, and Cosine Similarity. *G-Tech: Jurnal Teknologi Terapan*, 9(4), 2311–2321. <https://doi.org/10.70609/g-tech.v9i4.8306>
- Soleman, S. (2021). Pemanfaatan Metode Klasifikasi Naïve Bayes Untuk Pendekripsi Berita Hoax Pada Artikel Berbahasa Indonesia. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(2), 83. <https://doi.org/10.24014/coreit.v7i2.14290>
- Suciska, W., Eka, &, & Gunawibawa, Y. (2020). Pola Konsumsi Berita pada Kelompok Khalayak Digital di Kota Bandar Lampung. *Jurnal Ilmu Komunikasi*, 17(2), 249–265.
- Tarigan, A., Sahputra, I., & Multazam, T. (2024). SISFO : Jurnal Ilmiah Sistem Informasi Analisis Algoritma Naïve Bayes Classifier Untuk Mendekripsi Berita Hoax Pada Dinas Kominikasi Informatika Dan Persandian. *SISFO : Jurnal Ilmiah Sistem Informasi*, 8(1).
- Wibowo, A. R., Astuti, Y. P., Kartikadarma, E., Subhiyakto, E. R., Anisa, N., Winarsih, S., & Rohman, M. S. (2024). Penerapan Metode Naïve Bayes Classifier untuk Klasifikasi Sentimen pada Judul Berita. *Journal of Artificial Intelligence and Multimedia in Informatics*, 1(1), 1–12.
- Yoga Pratama, A., Ananda Sanjaya, G., Khairunisa Lubis, N., & Rangga Aditya, M. (2025). Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial. *METIK JURNAL*, 9, 92–99. <https://doi.org/10.47002/metik.v9i1.1055>
- Zufria, I., Lubis, A. H., & Febiyaula, S. S. (2024). ANALISIS SENTIMEN KEPERCAYAAN MASYARAKAT TERHADAP KEPOLISIAN REPUBLIK INDONESIA MENGGUNAKAN ALGORITMA SVM. *Journal of Science and Social Research*, 7(3), 1266–1272. <http://jurnal.goretanpena.com/index.php/JSSR>