

---

**ANALISIS SENTIMEN KINERJA KEPEMIMPINAN BUPATI DARI DATA KOMENTAR  
MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER  
(Studi : Kabupaten Lamongan)****M. Ali Na'im<sup>1</sup> dan Siti Mujilahwati<sup>2\*</sup>**

<sup>1</sup>Jurusan Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan  
Jl. Veteran No.53A, Jetis, Kec. Lamongan, Kabupaten Lamongan, Jawa Timur 62211  
Email: im.alina931@gmail.com<sup>1</sup>, moedjee@unisla.ac.id<sup>2</sup>

**Abstrak**

*Pada tingkatan Kabupaten, Bupati adalah kepala daerah yang memiliki peran penting. Bupati memiliki tanggung jawab yang mencakup penyelenggaraan pemerintahan daerah, termasuk dengan dikeluarkannya berbagai kebijakan yang ditujukan untuk kemajuan dan kesejahteraan masyarakat. Namun, kebijakan tersebut mendapat beragam tanggapan dan opini dari masyarakat. Sebagian masyarakat mungkin memberikan pendapat yang baik dan konstruktif maupun sebaliknya. Analisis sentimen yang dilakukan menggunakan metode TF-IDF, text mining serta algoritma Naïve bayes yang merupakan metode klasifikasi statistik yang berdasarkan pada Teorema Bayes. Metode Naïve bayes Classifier terdiri dari dua tahapan yaitu : Tahap pelatihan dilakukan proses analisis terhadap dokumen sampel berupa pemilihan kosakata yang merupakan kata-kata yang mungkin muncul dalam kumpulan dokumen serta menentukan probabilitas untuk setiap kategori. Hasil pada penelitian ini yang menggunakan sejumlah 200 data, yang terdiri dari 78 data yang bernilai positif serta 122 data yang bernilai negatif menunjukkan bahwa metode naïve bayes mampu mengklasifikasikan sentiment dengan nilai akurasi yang didapatkan sebesar 82%. Hasil perhitungan rata-rata precision sebesar 89%, perhitungan rata-rata Recall 77% serta perhitungan rata-rata F1-Score sejumlah 79%.*

**Kata kunci:** Analisis Sentimen, Bupati, Lamongan, Naïve bayes, TF-IDF

**1. PETUNJUK UMUM**

Pada tingkatan Kabupaten, Bupati adalah kepala daerah yang memiliki peran penting. Bupati memiliki tanggung jawab yang mencakup penyelenggaraan pemerintahan daerah, tugas dan wewenang Bupati berkaitan dengan berbagai aspek pemerintahan di tingkat Kabupaten, termasuk pengambilan kebijakan dan pelaksanaan program-program untuk masyarakat setempat (Dhuhri et al., 2021). Dalam kepemimpinan Bupati Kabupaten Lamongan, telah dikeluarkan berbagai kebijakan pada bidang pembangunan yang diarahkan untuk kemajuan dan kesejahteraan masyarakat, khususnya di Kecamatan Laren, kebijakan tersebut mendapat beragam tanggapan dan opini dari masyarakat. Tanggapan terhadap kebijakan tersebut dapat bervariasi. Sebagian masyarakat mungkin memberikan pendapat yang baik dan konstruktif mengenai kebijakan yang diambil, mengakui manfaatnya, dan mendukung upaya untuk memajukan pembangunan. Namun, tak jarang pula yang berpendapat kurang puas dengan pelayanan atau hasil kebijakan pada bidang pembangunan yang diberikan. Pendapat ini bisa berasal dari ekspektasi yang berbeda-beda atau pengalaman pribadi yang berbeda. Reaksi masyarakat, baik positif maupun konstruktif, merupakan bagian alami dari proses demokrasi dan kepemimpinan. Ini juga dapat menjadi masukan berharga bagi pemerintah untuk terus memperbaiki dan mengembangkan kebijakan yang lebih baik dan lebih sesuai dengan kebutuhan serta harapan masyarakat .

Analisis sentimen adalah metode yang digunakan untuk memahami dan mengolah data teks secara otomatis dengan tujuan mendapatkan informasi tentang sentimen atau perasaan yang tersirat dalam sebuah opini (Atthahara & Priyanti, 2021). Dalam konteks ini, analisis sentimen dapat diterapkan untuk memahami bagaimana masyarakat merasakan dan mengungkapkan pandangannya terhadap kinerja Bupati Lamongan selama periode 2021-2024 pada bidang pembangunan di Kecamatan Laren. Proses analisis sentimen melibatkan beberapa langkah, termasuk pengumpulan data teks yang berkaitan dengan kinerja pada bidang pembangunan khusus daerah Lamongan, pengolahan teks untuk mengidentifikasi kata-kata kunci dan makna, serta penentuan polaritas sentimen dari teks tersebut. Dengan menganalisis sentimen dari data teks yang dikumpulkan, dapat mendapatkan wawasan tentang bagaimana masyarakat Kecamatan Laren merasa puas terhadap kinerja Bupati pada bidang pembangunan .

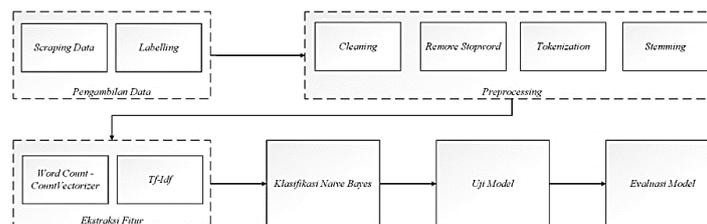
Naive Bayes adalah teknik text mining untuk membangun model pengklasifikasi yang sederhana tetapi memiliki keakuratan yang tinggi dalam mengklasifikasi (Muhammad Fadli Asshiddiqi, 2020; Mujilahwati, 2016). Untuk beberapa jenis model probabilitas, Naive Bayes Classifier dapat dilatih dengan sangat efisien dalam pengaturan supervised learning. Kelebihan dari Naive Bayes Classifier adalah metode ini hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter yang diperlukan dalam klasifikasi. Klasifikasi Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat dipublikasikan kedalam basis data dengan jumlah yang besar. Namun, penting untuk diingat bahwa analisis sentimen tidak selalu akurat 100% dan mungkin juga memerlukan interpretasi lebih lanjut. Pengumpulan data yang representatif dan perhatian terhadap konteks dan nuansa dalam teks juga penting dalam menjalankan analisis sentimen yang lebih efektif dan akurat (Ramadhani & Fajarianto, 2020).

Berdasarkan latar belakang tersebut, maka diharapkan penelitian analisis sentimen terhadap kepemimpinan Bupati Kabupaten Lamongan pada bidang pembangunan di Kecamatan Laren mampu mengelompokkan sentimen yang terjadi pada masyarakat terkait kepuasan kepemimpinan beliau, nantinya komentar masyarakat tersebut akan dibagi menjadi dua kelas yakni positif dan negatif menggunakan metode Naive bayes sehingga akan membantu para pemangku kebijakan dalam melakukan evaluasi terhadap program yang telah terealisasi dengan baik.

## 2. METODOLOGI

### 2.1. Deskripsi Alur Penelitian

Penelitian ini menggunakan metode naive bayes untuk melakukan analisis sentimen terhadap kinerja kepemimpinan Bupati Kabupaten Lamongan pada bidang pembangunan di Kecamatan Laren berdasarkan komentar masyarakat. Diagram alur penelitian ditampilkan pada gambar 1.



Gambar 1. Alur penelitian

### 2.2. Pengumpulan Data

Tahap awal penelitian, peneliti melakukan pengumpulan data menggunakan *Google Form* dengan responden yang berasal dari Kecamatan Laren Kabupaten Lamongan. Data yang telah terkumpul disimpan dengan format excel kemudian dilakukan penghapusan kolom yang tidak digunakan serta pelabelan kelas pada setiap data. Adapun kelas sentimen positif digantikan dengan label 0, sementara kelas sentimen negatif digantikan dengan label 1.

### 2.3. Preprocessing

Tujuan dari tahap *Preprocessing* adalah untuk menghilangkan elemen-elemen yang tidak diperlukan atau kata-kata yang tidak memberikan makna yang relevan. Proses ini melibatkan serangkaian langkah yang disesuaikan dengan karakteristik data komentar (Mujilahwati, 2016). Melalui tahap *Preprocessing*, diharapkan data menjadi lebih bersih dan terstruktur, sehingga dapat meningkatkan akurasi analisis yang akan dilakukan. Tahap *Pre-processing* ini melibatkan beberapa langkah seperti pembersihan (cleansing) data, tokenisasi (memecah teks menjadi kata-kata atau bagian), penyaringan (*filtering/remove stopword*), dan *stemming* (penghapusan imbuhan dalam kata) (Najiyah, 2023).

### 2.4. Ekstraksi Fitur

Dalam tahap pembuatan fitur, terdapat dua proses yang dilakukan: pembuatan word vector dan pembobotan kata dengan metode Tf-Idf. Word vector merujuk pada konversi fitur teks menjadi representasi vector dengan mengubah setiap kalimat yang telah diolah menjadi sejumlah array, yang kemudian menjadi matriks. Setiap baris dalam matriks ini mewakili sebuah dokumen, sedangkan kolom matriks merepresentasikan seluruh kata yang ada dalam seluruh teks yang terkumpul (Atika

et al., 2022). Selanjutnya memberi bobot pada kata dalam setiap kalimat menggunakan TF-IDF. Dengan pembobotan ini, dataset siap digunakan dalam tahap pelatihan dengan memanfaatkan perhitungan dalam algoritma Naïve bayes (Jimly Hanif et al., 2023). Berikut persamaannya:

$$tf = 0,5 + 0,5 * \frac{ft, d}{\max(ft, d)} \tag{1}$$

$$idf = \log \frac{N}{df} \tag{2}$$

$$W = tf * idf \tag{3}$$

Keterangan :

d	: Dokumen	N	: Total jumlah dokumen
t	: Kata pada dokumen	dtf	: Banyak dokumen yang mengandung kata i
ft,d	: Frekuensi kata pada d	idf	: Inversed Document Frequency
tf	: Banyaknya kata i pada sebuah dokumen	W	: Bobot dokumen ke-d terhadap kata ke t

### 2.5. Klasifikasi Naïve Bayes

Algoritma Naïve Bayes adalah metode klasifikasi statistik yang berdasarkan pada Teorema Bayes. Naïve Bayes adalah suatu algoritma pengklasifikasi yang menggunakan probabilitas dan statistik yang awalnya diusulkan oleh ilmuwan Inggris bernama Thomas Bayes (Hasri & Alita, 2022). Teorema Bayes digunakan untuk memprediksi peluang masa depan berdasarkan pengalaman sebelumnya. Dalam klasifikasi Naïve Bayes, proses lebih berfokus pada perhitungan dan estimasi probabilitas. Algoritma ini cocok digunakan dalam berbagai jenis klasifikasi, seperti klasifikasi teks, klasifikasi spam, dan banyak lagi (Azhar et al., 2022). Persamaannya adalah sebagai berikut:

$$P(H|X) = \frac{P(H|X).P(H)}{P(X)} \tag{4}$$

Keterangan :

X	: Data dengan class yang belum diketahui	P(X)	: Probabilitas X
H	: Hipotesis data X merupakan class spesifik	P(H)	: Probabilitas hipotesis H (prior probability)
P(H X)	: Probabilitas hipotesis H berdasar kondisi	P(X H)	: Probabilitas X berdasar kondisi hipotesis

### 2.6. Uji Model

Uji model dilakukan untuk mengevaluasi kinerja model yang telah dilatih. Sejumlah data diambil dari data pelatihan, biasanya sekitar 20% dari total data pelatihan. Proses pengambilan data dilakukan secara acak menggunakan pustaka atau library Python yang sesuai. Setelah uji model selesai, hasilnya akan menunjukkan seberapa besar akurasi dari metode yang diterapkan pada data.

### 2.7. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja model yang dibuat menggunakan *confusion matrix*. Confusion matrix dapat digunakan untuk menganalisa seberapa baik classifier mengenali tuple dari kelas yang berbeda. TP dan TN memberikan informasi ketika klasifikasi benar, sedangkan FP dan FN memberikan informasi ketika klasifikasi salah (Sihombing & Nurdiyanto, 2023). Setelah data pengujian dijalankan melalui model yang telah dihasilkan oleh data pelatihan, prediksi kelas akan dihasilkan dari data pengujian. Hal ini memungkinkan perhitungan nilai akurasi, presisi, *Recall*, dan *F1-Score* (Hasri & Alita, 2022). Berikut adalah persamaannya :

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - Score = 2 \times \frac{Precision * recall}{Precision + recall} \tag{8}$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

Data yang telah terkumpul kemudian diekspor dalam bentuk excel dan dilakukan pelabelan (0) untuk kelas positif dan (1) untuk kelas negatif. Adapun tampilan dataset sebagaimana tabel 1.

Tabel 1. Dataset

Text	Class
Kinerja lumayan bagus, namun jalan didesa desa masih banyak yang rusak, bantuan untuk rakyat miskin juga masih belum tepat sasaran. Intinya masyarakat didesa desa kecil harus lebih diperhatikan lagi.	0
Sekarang ini Masih bagus dan terlaksana sesuai yang ada pada anggaran	0
rate kinerja dalam hal pembangunan infrastruktur seperti halnya perbaikan jalan rusak masih jauh dari kata maksimal, kisaran 30%, terbukti dari banyaknya jalan rusak, bukan hanya jalanan desa saja, bahkan jalan raya pantura masih banyak ditemui jalan berlubang, kalau hujan air menggenang, tidak adanya saluran air yg mampu menampung air di jalanan sehabis hujan.	1
Kurang bisa dirasakan masyarakat kawasan Pantura Pak.	1
Ya seperti sebelumnya jalan masih banyak yang rusak. Apalagi didesa sinyal saja susah . Lebih ditingkatkan lagi kerjanya bapak	1

#### 3.2. Preprocessing

Data yang telah dilabeli kemudian masuk tahapan *preprocessing* yang terdiri dari proses *cleaning*, *remove stopwords*, *tokenize* dan *stemming*. Adapun hasil tahapan preprocessing pada data index [0] adalah sebagaimana tabel 2.

Tabel 2. Hasil tahapan preprocessing

Proses	Text	Result
<i>Cleaning</i>	Kinerja lumayan bagus, namun jalan didesa desa masih banyak yang rusak, bantuan untuk rakyat miskin juga masih belum tepat sasaran. Intinya masyarakat didesa desa kecil harus lebih diperhatikan lagi.	kinerja lumayan bagus namun jalan didesa desa masih banyak yang rusak bantuan untuk rakyat miskin juga masih belum tepat sasaran intinya masyarakat didesa desa kecil harus lebih diperhatikan lagi
<i>Remove Stopword</i>	kinerja lumayan bagus namun jalan didesa desa masih banyak yang rusak bantuan untuk rakyat miskin juga masih belum tepat sasaran intinya masyarakat didesa desa kecil harus lebih diperhatikan lagi	kinerja lumayan bagus jalan didesa desa rusak bantuan rakyat miskin sasaran intinya masyarakat didesa desa diperhatikan
<i>Tokenize</i>	kinerja lumayan bagus jalan didesa desa rusak bantuan rakyat miskin sasaran intinya masyarakat didesa desa diperhatikan	['kinerja', 'lumayan', 'bagus', 'jalan', 'didesa', 'desa', 'rusak', 'bantuan', 'rakyat', 'miskin', 'sasaran', 'intinya', 'masyarakat', 'didesa', 'desa', 'diperhatikan']
<i>Stemming</i>	['kinerja', 'lumayan', 'bagus', 'jalan', 'didesa', 'desa', 'rusak', 'bantuan', 'rakyat', 'miskin', 'sasaran', 'intinya', 'masyarakat', 'didesa', 'desa', 'diperhatikan']	kerja lumayan bagus jalan desa desa rusak bantu rakyat miskin sasar inti masyarakat desa desa perhati

#### 3.3. Ekstraksi Fitur

Proses yang dilakukan yaitu pembuatan *word vector* dan pembobotan kata dengan *Term Frequency-Inversed Document Frequency* (TF-IDF). Adapun hasil pembuatan *vector* kata adalah sebagaimana gambar 2.

```
In [8]: # jumlah kolom dan baris TF-IDF matrix consists
print("n_samples: %d, n_features: %d" % vectors.shape)
n_samples: 200, n_features: 509
```

Gambar 2. Word Vector

Representasi tersebut menghasilkan 200 data yang mempunyai 509 kata. Setelah proses tersebut selesai selanjutnya adalah proses perhitungan *TF-IDF* dari masing-masing kata. Adapun hasil dari perhitungan TF-IDF dari indeks [0] adalah sebagaimana tabel 3.

**Tabel 3. Perhitungan TF-IDF**

Term	TF	DF	IDF	W
kerja	0,0625	33	1,78	0,11
lumayan	0,0625	13	2,66	0,16
bagus	0,0625	35	1,71	0,10
jalan	0,0625	103	0,65	0,04
desa	0,25	42	1,54	0,38
rusak	0,0625	30	1,86	0,11
bantu	0,0625	1	4,60	0,28
rakyat	0,0625	2	4,19	0,26
miskin	0,0625	1	4,60	0,28
sasar	0,0625	3	3,91	0,24
inti	0,0625	1	4,60	0,28
masyarakat	0,0625	20	2,25	0,14
perhati	0,0625	21	2,20	0,13

### 3.4. Klasifikasi Naïve Bayes

Membangun model yang bakal digunakan untuk klasifikasi pada data uji. Dalam proses klasifikasi, digunakan data uji sebesar 20% dan data latih sebesar 80% dari seluruh data yang telah diproses sebelumnya. Adapun proses pembagiannya sebagaimana gambar 3.

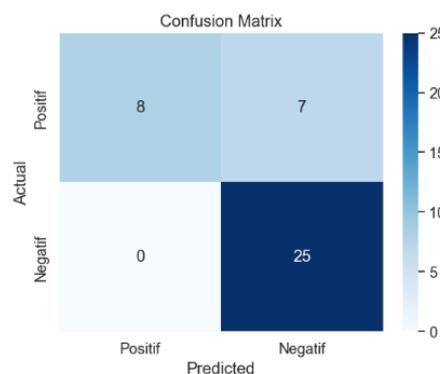
```
In [55]: print("n_samples: %d, n_features: %d" % tfidf_train_vectors.shape)
print("n_samples: %d, n_features: %d" % tfidf_test_vectors.shape)
n_samples: 160, n_features: 447
n_samples: 40, n_features: 447
```

**Gambar 3. Pembagian data train dan data test**

Setelah seluruh tahapan berhasil diselesaikan, langkah berikutnya adalah menghitung kinerja dari algoritma yang digunakan untuk evaluasi.

### 3.5. Evaluasi Model

Pada gambar 3. dijelaskan bahwa dari 200 dataset dilakukan pembagian 80% untuk data latih dan 20% untuk data uji. Adapun jumlah untuk data latih sebanyak 160 data dan 447 kata sedangkan untuk data uji sebanyak 40 data dan 447 kata. Adapun tampilan matriks konfusi dari hasil klasifikasi adalah sebagaimana gambar 4.



**Gambar 4. Matriks konfusi**

Terdapat 8 True Positives (TP) dan 25 True Negatif (TN), sementara terdapat 0 False Positives (FP) dan 7 False Negatif (FN). Langkah selanjutnya adalah perhitungan akurasi *naïve bayes*, presisi, *recall* dan *f1-score*. Hasil perhitungannya dapat dilihat pada tabel 4.

**Tabel 4. Perhitungan akurasi, presisi, recall dan f1-score**

Jenis	Presisi	Recall	F1-Score	Akurasi
Positif	100%	53%	70%	82%
Negatif	78%	100%	88%	
Rata-rata	89%	77%	79%	

#### 4. KESIMPULAN

Berdasarkan proses yang telah dilakukan, alur melakukan analisis sentimen pada penelitian ini dengan tahapan pengumpulan data, pelabelan data, *preprocessing*, ekstraksi fitur *TF-IDF*, pengklasifikasian dengan metode *naïve bayes* dan uji model. Hasil akurasi yang didapatkan pada penelitian ini dengan menggunakan metode *naïve bayes* adalah 82%. Dengan perincian nilai presisi kelas positif 100% dan negatif 78% dengan nilai rata-rata presisi 89%, nilai recall kelas positif 53% dan negatif 100% dengan nilai rata-rata *recall* 77% serta nilai f1-score kelas positif 70% dan nilai f1-score 88% dengan nilai rata-rata *f1-score* 79%.

#### DAFTAR PUSTAKA

- Atika, D., Ari Aldino, A., Informasi, S., Pagar Alam No, J., Ratu, L., & Kedaton, K. (2022). Term Frequency-Inverse Document Frequency Support Vector Machine Untuk Analisis Sentimen Opini Masyarakat Terhadap Tekanan Mental Pada Media Sosial Twitter. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 3(4), page-page.
- Atthahara, H., & Priyanti, E. (2021). Perempuan Kepala Daerah: Analisis Kinerja Bupati Cellica Nurrachadiana dalam Program Kepentingan Praktis Gender di Kabupaten Karawang. *Jurnal Politikom Indonesiana*, 4(2), 72–86. <https://doi.org/10.35706/jpi.v4i2.3234>
- Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(1), 267–281.
- Dhuhri, A., Isnaini, I., & Lubis, M. S. (2021). Analisis Kebijakan Peraturan Bupati No. 6 Tahun 2019 Di Lingkungan Kabupaten Tapanuli Tengah Dalam Meningkatkan Kinerja Pegawai Di Kelurahan Sibuluan Indah. *Perspektif*, 11(1), 124–130. <https://doi.org/10.31289/perspektif.v11i1.5363>
- Hasri, C. F., & Alita, D. (2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal Informatika Dan Rekayasa Perangkat Lunak (JATIKA)*, 3(2), 145–160.
- Jimly Hanif, A., Nur Farid, M., & Hasanah, B. (2023). Penerapan Natural Language Processing untuk Klasifikasi Bidang Minat berdasarkan Judul Tugas Akhir. *Jurnal Sistim Informasi Dan Teknologi*, 2(1), 41–49. <https://doi.org/10.37034/jsisfotek.v5i1.196>
- Muhammad Fadli Asshiddiqi, K. M. L. (2020). *Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI*. 39(6), 0–2.
- Mujilawati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 2089–9815.
- Najiyah, I. (2023). Analisis Sentimen Tanggapan Masyarakat Indonesia Tentang Kenaikan Bbm Menggunakan Metode Artificial Neural Network. *Jurnal Responsif*, 5(1), 92–100.
- Ramadhani, N., & Fajarianto, N. (2020). Sistem Informasi Evaluasi Perkuliahan dengan Sentimen Analisis Menggunakan Naïve Bayes dan Smoothing Laplace. *Jurnal Sistem Informasi Bisnis*, 10(2), 228–234. <https://doi.org/10.21456/vol10iss2pp228-234>
- Sihombing, C. P., & Nurdianto, H. (2023). *Data mining untuk classification dalam mendeteksi kerusakan elektrokardiogram menggunakan naive bayes*. 8(3), 792–801.