

## **Analisis Sentimen Pada Pelayanan Publik DPMPTSP Kabupaten Magelang Menggunakan Naïve Bayes**

**Istin Nuryati<sup>1,\*</sup>, Purwono Hendradi<sup>1</sup>, Ardhin Primadewi<sup>2</sup>**

<sup>1</sup> Fakultas Teknik, Teknik Informatika, Universitas Muhammadiyah Magelang, Magelang, Indonesia

<sup>2</sup> Fakultas Teknik, Teknik Informatika, Universitas Muhammadiyah Magelang, Magelang, Indonesia

Email: <sup>1,\*</sup>istinnuryati2025@gmail.com, <sup>2</sup>p\_hendra@ummg.ac.id., <sup>3</sup>ardhin@ummg.ac.id

Email Penulis Korespondensi: istinnuryati2025@gmail.com

**Abstrak**—Penelitian ini bertujuan melakukan analisis sentimen terhadap pelayanan publik di DPMPTSP Kabupaten Magelang menggunakan metode *Naïve Bayes* pada data saran masyarakat dari Survei Kepuasan Masyarakat (SKM) Triwulan I sampai IV Tahun 2024. Data berupa teks komentar saran yang bersifat subjektif diolah menggunakan teknik text mining mulai dari pembersihan data, normalisasi, tokenisasi, *stopword removal*, hingga *stemming*. Selanjutnya, pelabelan sentimen dilakukan dengan dua metode, yaitu menggunakan *InSet Lexicon* dan model *IndoBERT*. Setelah itu, fitur teks diberi bobot dengan metode TF-IDF dan dilakukan klasifikasi sentimen positif dan negatif dengan *Naïve Bayes*.

Hasil penelitian menunjukkan bahwa mayoritas saran masyarakat mengandung sentimen positif. Model *Naïve Bayes* dengan pelabelan *InSet Lexicon* menghasilkan akurasi 89,13%, dengan keseimbangan yang lebih baik antara kelas positif dan negatif dibandingkan pelabelan *IndoBERT* yang menghasilkan akurasi 90,22% namun kurang seimbang terutama pada kelas negatif. Visualisasi *wordcloud* dari saran positif menunjukkan kata-kata dominan yang mengarah pada apresiasi layanan cepat, ramah, dan mudah, sementara saran negatif mengandung kata-kata terkait perbaikan pada respons pengaduan dan waktu penyelesaian.

Penelitian ini menyimpulkan bahwa analisis sentimen otomatis menggunakan *Naïve Bayes* efektif untuk menganalisa saran masyarakat pada pelayanan publik dan dapat menjadi bahan evaluasi untuk peningkatan kualitas layanan DPMPTSP Kabupaten Magelang.

**Kata Kunci:** Analisis Sentimen, *Naïve Bayes*, Pelayanan Publik, Survei Kepuasan Masyarakat (SKM), *Text Mining*, TF-IDF, *InSet Lexicon*, *IndoBERT*, Klasifikasi Teks

**Abstract**—This research aims to conduct sentiment analysis on public services at the DPMPTSP of Magelang Regency using the *Naïve Bayes* method on community suggestion data from the Public Satisfaction Survey (SKM) from the first to the fourth quarter of 2024. The data, consisting of subjective text comments, are processed using text mining techniques including data cleaning, normalization, tokenization, *stopword removal*, and *stemming*. Next, sentiment labeling is carried out using two methods: the *InSet Lexicon* and the *IndoBERT* model. Afterwards, the text features are weighted using the TF-IDF method, and classification of positive and negative sentiments is performed using *Naïve Bayes*.

The results indicate that the majority of community suggestions contain positive sentiments. The *Naïve Bayes* model with *InSet Lexicon* labeling achieves an accuracy of 89.13%, showing a more balanced classification between positive and negative classes compared to *IndoBERT* labeling, which obtains an accuracy of 90.22% but is less balanced, especially in the negative class. Word cloud visualization of positive suggestions highlights dominant words related to appreciation for fast, friendly, and easy service, whereas negative suggestions include words related to improvements in complaint response and processing time.

This study concludes that automatic sentiment analysis using *Naïve Bayes* is effective for mapping community opinions on public services and can serve as a basis for evaluating and improving the quality of services at DPMPTSP Magelang Regency.

**Keywords:** Sentiment Analysis, *Naïve Bayes*, Public Service, Public Satisfaction Survey (SKM), *Text Mining*, TF-IDF, *InSet Lexicon*, *IndoBERT*, Text Classification

## **1. PENDAHULUAN**

Undang-Undang Nomor 25 Tahun 2009[1] dan Peraturan Pemerintah Nomor 96 Tahun 2012 [2] mengamanatkan keterlibatan masyarakat dalam penyelenggaraan pelayanan publik guna mewujudkan sistem yang adil, transparan, dan akuntabel. Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Magelang [3] bertugas memberikan pelayanan publik di bidang perizinan sesuai Peraturan Bupati Nomor 19 Tahun 2023[4], dengan tujuan meningkatkan kepastian hukum, perlindungan masyarakat, kemudahan berusaha, dan iklim investasi di daerah. Salah satu indikator kinerja utama DPMPTSP adalah Nilai Survei Kepuasan Masyarakat (SKM)[5] yang digunakan sebagai dasar evaluasi dan perbaikan pelayanan publik.

Hasil SKM periode 2020–2024 menunjukkan capaian nilai yang bervariasi, dengan skor tertinggi pada aspek perilaku dan kompetensi pelaksana, serta aspek persyaratan pelayanan, sistem, dan waktu penyelesaian sebagai yang terendah. Namun, saran dari masyarakat dalam SKM belum diolah secara komprehensif untuk mengidentifikasi tren permasalahan.

Analisis sentimen merupakan teknik *text mining* [6] untuk mengklasifikasikan opini menjadi sentimen positif, negatif, atau netral secara otomatis [7]. Penelitian sebelumnya menggunakan metode *Support Vector Machine* (SVM) dan *Multinomial Naïve Bayes* (MNB) [8] dengan pelabelan *InSet Lexicon* dalam analisis sentimen pelayanan publik Pemerintah DKI Jakarta, menunjukkan akurasi SVM sebesar 82% lebih tinggi dari MNB [9]. Penelitian lain menganalisis sentimen masyarakat di media sosial Twitter terkait kebijakan Kominfo dengan metode *K-Nearest Neighbors* (KNN) dan *Random Forest* yang mencapai akurasi tertinggi 85,8%, sedangkan *Decision Tree* memiliki akurasi terendah 83,8% [10]. Sementara itu, *Naïve Bayes* [11] dianggap efektif untuk klasifikasi teks pendek seperti saran SKM karena



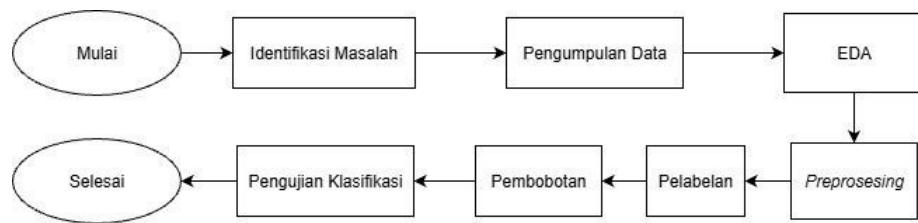
memanfaatkan probabilitas kemunculan kata dan asumsi independen antar fitur [12]. Analisis sentimen ini memudahkan pemahaman pola kepuasan masyarakat secara luas tanpa harus membaca teks satu per satu [13]. Data sentimen diolah melalui *preprocessing* [14] dan *pembobotan TF-IDF* [15], kemudian diklasifikasikan menjadi sentimen positif dan negatif [16] menggunakan pelabelan *InSet Lexicon* [17] dan *IndoBert* [18] dengan bahasa pemrograman *Python* [19] dan memanfaatkan *library* [20] pada *Python*.

Penelitian ini bertujuan menerapkan metode pelabelan *InSet Lexicon* dan *IndoBERT* dalam analisis sentimen pelayanan publik DPMPTSP Kabupaten Magelang serta melakukan klasifikasi sentimen menggunakan Naïve Bayes. Hasil analisis sentimen akan divisualisasikan dan diharapkan menjadi referensi evaluasi dan peningkatan kualitas pelayanan publik. Rumusan masalah penelitian ini adalah: (1) bagaimana penerapan pelabelan *InSet Lexicon* dan *IndoBERT* dalam analisis sentimen pelayanan publik, dan (2) bagaimana klasifikasi sentimen saran SKM menggunakan Naïve Bayes. Manfaat penelitian ini meliputi kontribusi teoritis dalam pengembangan machine learning untuk analisis sentimen dan manfaat praktis berupa peningkatan kualitas pelayanan publik di DPMPTSP Kabupaten Magelang melalui hasil evaluasi yang akurat dan berbasis data.

## 2. METODOLOGI PENELITIAN

### 2.1 Metode Penelitian

Penelitian ini menggunakan tahapan 7 tahapan yang dapat dilihat sesuai gambar 1.



**Gambar 1.** Tahapan penelitian

Tahapan yang dilakukan yaitu identifikasi masalah, pengumpulan data, *Exploratory Data Analysys (EDA)*, preprocessing, pelabelan, pembobotan dan pengujian klasifikasi. Penelitian diawali dengan proses identifikasi masalah, dilanjutkan pengumpulan data dimana data yang digunakan adalah data primer berupa hasil Survei Kepuasan Masyarakat Triwulan I sd Triwulan IV Tahun 2024. Langkah selanjutnya yaitu *Exploratory Data Analysys (EDA)* yang merupakan tahap awal dalam analisis data yang berfokus pada proses mengeksplorasi, memahami, serta memperoleh *insight* awal dari data sebelum dilakukan pemodelan atau penarikan kesimpulan. Dilanjutkan dengan langkah *preprocessing* pada tahap ini dilakukan pembersihan data, normalisasi kata, *tokenize*, *stopwords* dan *stemming*. Data hasil preprocessing kemudian dilakukan pelabelan menggunakan *InSet Lexicon* dan *IndoBert* untuk mengklasifikasikan sentimen ke dalam kelas negatif dan positif. Hasil pelabelan selanjunya digunakan untuk melatih data untuk model klasifikasi *Naïve Bayes* yang dilakukan pengujian klasifikasi metode naïve bayes menggunakan *Confussion Matrix* untuk mengetahui indicator kinerja *accuracy*, *precision*, *recall*, dan *F1-score*.

### 2.2 Hipotesis

Berdasarkan tujuan dan rumusan masalah yang telah ditetapkan, maka hipotesis dalam penelitian ini dirumuskan sebagai berikut:

1. Hipotesis 0 (H0): Penggunaan pelabelan *InSet Lexicon* dan *IndoBERT* tidak berpengaruh terhadap akurasi analisis sentimen.
2. Hipotesis 1 (H1): Penggunaan pelabelan *InSet Lexicon* dan *IndoBERT* berpengaruh terhadap akurasi analisis sentimen.

### 2.3 Simulasi

Tahapan simulasi merupakan tahapan rancangan percobaan. Dalam tahap ini dilakukan percobaan dengan menggunakan sampel data sebanyak 3 (tiga) data yang ditunjukkan pada tabel 1.

**Tabel 1.** Sampel Data

No	Sampel data
1	tracking perizinan, mudah, notifikasi, rinci
2	pelayanan, lama, waktu, estimasi
3	pelayanan, perizinan, baik

#### 2.3.1 Pelabelan

##### 1. Pelabelan *InSet Lexicon*

Klasifikasi model *Naïve Bayes* dengan pelabelan *InSet Lexicon* menggunakan kamus berbahasa Indonesia yaitu kamus *InSetPos* yang memiliki 3.609 entri positif dan *InSetNeg* yang memiliki 6.609 entri negatif yang

masing-masing memiliki skor dalam rentang -5 hingga +5. Setiap data sampel kemudian dipisah dalam bentuk token/kata. Setiap kata pada data sampel kemudian dihitung berdasarkan skor pada kamus *InSetPos* dan kamus *InSetNeg*, apabila ada kata yang tidak terdapat pada kamus maka diberi skor nol (0). Setiap kalimat pada data sampel kemudian dihitung total skor polaritasnya dengan cara menjumlahkan skor masing-masing kata berdasarkan skor pada kamus *InSetPos* dan kamus *InSetNeg*. Perhitungan total skor pada data sampel sebagai berikut :

**Tabel 2.** Perhitungan skor polaritas

No	Sampel data	Perhitungan skor polarity	Total Skor Polaritas
1	tracking, perizinan, mudah, notifikasi, rinci	0+0+1+0+0	1
2	pelayanan, lama, waktu, estimasi	0+(-1)+0+0	-1
3	pelayanan, perizinan, baik	0+0+1	1

Kemudian untuk menentukan kelas positif dan negatif ditentukan berdasarkan total skor polaritas dimana total skor polaritas kurang dari nol ( $<0$ ) masuk ke dalam kelas negatif, sedangkan total skor polaritas sama dengan atau lebih dari nol ( $\geq 0$ ) masuk ke dalam kelas positif. Hasil perhitungan total skor polaritas terhadap sampel data dapat dilihat pada tabel 3.

**Tabel 3.** Hasil pelabelan *InSet Lexicon*

No	Sampel data	Total Skor Polaritas	Label
1	tracking, perizinan, mudah, notifikasi, rinci	2	Positif
2	pelayanan, lama, waktu, estimasi	-1	Negatif
3	pelayanan, perizinan, baik	1	Positif

## 2. Pelabelan *IndoBert*

Pelabelan *IndoBert* diawali dengan langkah tokenisasi. Pada tahap ini teks diubah ke token sesuai dengan *vocabulary BERT*, dalam penelitian ini model *IndoBERT* yang digunakan yaitu *RoBERTa-base Indoneisan* yang memiliki lebih dari  $\pm 30.000$  token *vocabulary*. Token tersebut kemudian dikonversi menjadi ID angka oleh *tokenizer IndoBERT*. Token ID kemudian dimasukkan ke dalam model sehingga menghasilkan logit (output mentah) untuk tiap kelas. Dalam hal ini metode pelabelan *IndoBERT* secara otomatis mempunyai 3 kelas, positif, netral dan negatif. Output logit dari setiap data sampel ditunjukkan pada tabel 4.

**Tabel 4.** Hasil perhitungan logit

No	Sampel data	Logit Positive	Logit Neutral	Logit Negative
1	tracking, perizinan, mudah, notifikasi, rinci	2,1	0,3	-1,0
2	pelayanan, lama, waktu, estimasi	-0,5	-0,2	1,2
3	pelayanan, perizinan, baik	1	0,1	-0,8

Setelah itu dilakukan perhitungan  $e^{\text{logit}}$  dari masing-masing dokumen beserta total  $e^{\text{logit}}$  yang hasilnya ditunjukkan pada tabel 5.

**Tabel 5.** Hasil perhitungan  $e^{\text{logit}}$ 

No	Sampel data	$e^{\text{Positive}}$	$e^{\text{Neutral}}$	$e^{\text{Negatif}}$	Total $e^{\text{logit}}$
1	tracking, perizinan, mudah, notifikasi, rinci	8,1662	1,3499	0,3679	9,884
2	pelayanan, lama, waktu, estimasi	0,6065	0,8187	3,3201	4,7453
3	pelayanan, perizinan, baik	6,0496	1,1052	0,4493	7,6041

Kemudian dilakukan perhitungan *probabilitas softmax* dengan menggunakan rumus sebagai berikut :

$$P_i = \frac{e^{\text{logit}_i}}{\sum_j e^{\text{logit}_j}}$$

Atau dengan kata lain dalam perhitungan probabilitas softmax masing-masing kelas sebagai berikut :

$$P(\text{Positive}) = e^{\text{Positive}} / \text{Total}$$

$$P(\text{Neutral}) = e^{\text{Neutral}} / \text{Total}$$

$$P(\text{Negative}) = e^{\text{Negative}} / \text{Total}$$

Hasil perhitungan *probabilitas softmax* dapat dilihat pada tabel 6.

**Tabel 6.** Hasil perhitungan *probabilitas softmax IndoBert*

No	Sampel data	P(Positive)	P(Neutral)	P(Negatif)	Label
1	tracking, perizinan, mudah, notifikasi, rinci	8,1662	1,3499	0,3679	Positive
2	pelayanan, lama, waktu, estimasi	0,6065	0,8187	3,3201	Negative
3	pelayanan, perizinan, baik	6,0496	1,1052	0,4493	Positive



Pelabelan pada *IndoBert* ditentukan oleh nilai *probabilitas softmax* tertinggi, sebagai contoh pada dokumen 1 nilai pada  $P(Positive)$  sehingga dokumen tersebut diberikan label *positive*. Pada penelitian ini hanya menggunakan dua kelas pelabelan yaitu *positive* dan *negative*, sehingga apabila nilai probabilitas tertinggi pada suatu dokumen terdapat pada kelas *neutral*, maka di mapping ke kelas *positive*.

### 2.3.2 Pembobotan TF-IDF

Pada tahapan ini dilakukan percobaan perhitungan manual *Term Frequency-Inverse Document Frequency (TF-IDF)* menggunakan aplikasi Microsoft excel. Hasil *TF-IDF* seperti terlihat pada tabel 7.

**Tabel 7.** Perhitungan TF-IDF

Term	TF-IDF		
	d1 (positif)	d2 (negatif)	d3 (positif)
Tracking	0,095	0	0
Perizinan	0,035	0	0,059
Mudah	0,095	0	0
Notifikasi	0,095	0	0
Rinci	0,095	0	0
Pelayanan	0	0,044	0,059
Lama	0	0,119	0
Waktu	0	0,119	0
Estimasi	0	0,119	0
Baik	0	0	0,159

### 2.3.3 Klasifikasi sentimen menggunakan *Naïve Bayes*

Setelah dilakukan pembobotan dengan TF-IDF kemudian dilakukan klasifikasi menggunakan metode *Naïve Bayes* dengan langkah-langkah berikut :

1. Menghitung Probabilitas (P|Kelas)

Probabilitas (P|Kelas) dari sejumlah tiga (3) sampel data adalah sebagai berikut :

$$P(\text{Positif}) = 2/3 = 0,6667$$

$$P(\text{Negatif}) = 1/3 = 0,333$$

2. Menghitung Total TF-IDF per Kelas

a. Positif (d1+d3)

$$\text{Total}_{\text{positif}} = \text{Jumlah semua nilai di di } d1 = 0,095 + 0,035 + 0,095 + 0,095 + 0,095 + 0,059 + 0,059 + 0,159 = 0,692$$

b. Negatif (d2)

$$\text{Total}_{\text{negatif}} = \text{Jumlah semua nilai di di } d2 = 0,044 + 0,119 + 0,119 + 0,119 = 0,401$$

3. Menghitung *likelihood*

Menggunakan rumus sebagai berikut :

$$P(\text{term}_i | \text{kelas}) = \frac{\text{TF} - \text{IDF term}_i + 1}{\text{total TF} - \text{IDF semua term di kelas} + V}$$

Dimana V adalah jumlah term unik = 10

Contoh hasil perhitungan *likelihood* pada term “tracking” sebagai berikut :

$$P(\text{tracking} | \text{positif}) = \frac{0,095+1}{0,692+10} = 0,1024$$

$$P(\text{tracking} | \text{negatif}) = \frac{0+1}{0,401+10} = 0,0961$$

Hasil perhitungan *likelihood* secara lengkap dapat dilihat pada tabel 8.

**Tabel 8. Hasil Perhitungan likelihood**

Term	P(wi positif)	P(wi negatif)
Tracking	0,1024	0,0961
Perizinan	0,1023	0,0961
Mudah	0,1024	0,0961
Notifikasi	0,1024	0,0961
Rinci	0,1024	0,0961
Pelayanan	0,0990	0,1004
Lama	0,0935	0,1076
Waktu	0,0935	0,1076
Estimasi	0,0935	0,1076
Baik	0,1084	0,0961

#### 1. Menghitung *Posterior Probabilitas*

Perhitungan *posterior probabilitas* dilakukan dengan menggunakan rumus

$n$

$$P(\text{Kelas} | \text{Dokumen}) = P(\text{Kelas}) \times \prod_{i=1}^n P(\text{term}_i | \text{Kelas})^{TF_i}$$

Sebagai contoh dalam melakukan prediksi dokumen yang mengandung term [perizinan, mudah, waktu], akan menghasilkan prediksi seperti sebagai berikut :

$$\begin{aligned} P(\text{Positif} | \text{Dokumen}) &= P(\text{Positif}) \times P(\text{perizinan} | \text{Positif}) \times P(\text{mudah} | \text{positif}) \times P(\text{waktu} | \text{positif}) \\ &= 0,667 \times 0,1023 \times 0,1024 \times 0,0935 \\ &= 0,000653376 \end{aligned}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil

#### 3.1.1 Implementasi Metode

##### a) Identifikasi Masalah

DPMPTSP Kabupaten Magelang merupakan Organisasi Perangkat Daerah di Kabupaten Magelang yang menyelenggarakan pelayanan publik kepada masyarakat. DPMPTSP Kabupaten Magelang memiliki permasalahan terkait identifikasi komentar masyarakat terhadap pelayanan publik yang diberikan yang termuat dalam survei kepuasan masyarakat(SKM). Dalam hasil survei kepuasan masyarakat per triwulan dapat terlihat penilaian secara kuantitatif dari masyarakat atas unsur-unsur penilaian survei. Namun dalam rangka perbaikan pelayanan, DPMPTSP membutuhkan gambaran komentar yang jelas dan spesifik berdasarkan pengalaman nyata dari masyarakat yang termuat dalam saran pada SKM.

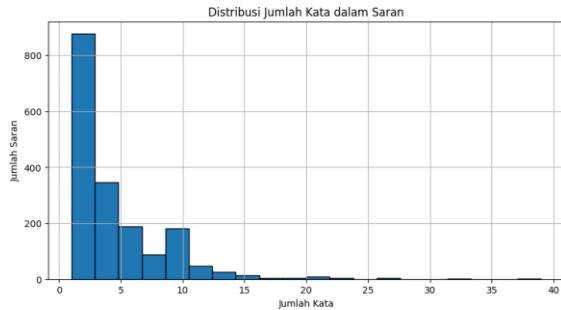
##### b) Pengumpulan Data

Data yang digunakan merupakan data primer dari DPMPTSP Kabupaten Magelang yaitu data hasil Survey Kepuasan Masyarakat Triwulan I sd Triwulan IV Tahun 2024 dengan jumlah 1.789 data yang didalamnya memuat kesan dan saran dari masyarakat atas pelayanan publik yang diperoleh. Dataset berupa file excel Laporan SKM triwulan I sd. IV yang kemudian diubah kedalam format csv dengan nama file skm2024.csv. Dataset yang digunakan terdiri dari kolom No, Umur, Jenis Kelamin, Pendidikan, Pekerjaan, U1, U2, U3, U4, U5, U6, U7, U8, U9, KESAN ATAS PELAYANAN YANG DITERIMA, SARAN UNTUK PERBAIKAN PELAYANAN. Namun fokus penelitian ini pada kolom SARAN UNTUK PERBAIKAN PELAYANAN karena kolom yang lain merupakan numerik yang tidak bisa dilakukan analisis sentimen.

##### c) Exploratory Data Analysis (EDA)

Pada tahap ini dilakukan analisis awal terhadap dataset. Hal yang dilakukan diantaranya mengecek jumlah kolom dan baris (*data.shape*), menampilkan tipe data(*data.dtypes*) dan melihat ringkasan statistik dari kolom pada DataFrame (*data.describe*). Setelah itu dilakukan seleksi kolom yang dilakukan analisis sentimen Pada tahap ini dilakukan seleksi kolom yang akan dilakukan analisis sentimen yaitu kolom “*SARAN UNTUK PERBAIKAN PELAYANAN*”. Kemudian dilanjutkan penghapusan data kosong (*data.dropna()*). Setelah itu dilakukan pengecekan untuk melihat distribusi jumlah kata dalam komentar. Distribusi jumlah kata dalam komentar dapat dilihat pada gambar 2.





**Gambar 2.** Distribusi Jumlah Kata dalam Saran

Dari gambar tersebut terlihat *dataset* didominasi dengan saran yang memiliki jumlah 1-5 kata. Saran yang hanya mengandung sedikit kata cenderung tidak memberikan informasi yang cukup untuk menentukan sentimen secara akurat. Sehingga dalam hal ini perlu dilakukan pembatasan jumlah kata pada saran yang akan dilakukan analisis sentimen adalah saran yang memiliki kata antara 3 sampai dengan 50 kata. Setelah dilakukan pembatasan jumlah kata dan setelah dilakukan penghapusan data kosong, dari *dataset* sejumlah 1789 data menghasilkan 912 data saran terpilih. Selanjutnya dilakukan visualisasi *WordCloud* dataset sebelum dilakukan *preprosesing*. Hal ini dilakukan untuk mengetahui kata yang paling banyak/sering muncul. Berikut *WordCloud* dataset awal sebelum *preprocessing*.



### Gambar 3. WordCloud Saram

Dari hasil *WordCloud* dapat dilihat bahwa terdapat kata umum seperti "untuk", "saat", dan "sudah" yang tidak memiliki makna signifikan terhadap sentimen, sehingga dikategorikan sebagai *stopwords* yang umumnya dihapus pada tahap preprocessing agar *WordCloud* menampilkan kata-kata yang lebih bermakna.

### a) Preprocessing

*Preprocessing* dilakukan pada dataset dari saran terpilih berjumlah 921 data saran, proses yang dilakukan pada *preprocessing* yaitu pembersihan data, normalisasi kata, *Stopwords*, *Tokenize*, dan *Stemming*.

## 1. Pembersihan data

Pada tahap pembersihan data dilakukan karakter lain selain a-z, simbol dan angka. Langkah ini memanfaatkan library python “*re*”. Hasil Pembersihan data ditunjukkan pada tabel 9.

**Tabel 9.** Hasil Pembersihan Data

Data awal	Hasil Pembersihan Data
Pelayanan cepat, belum ada 9 hari sudah selesai. Semoga tetap bisa memberikan pelayanan yg terbaik setiap harinya dengan sepenuh hati	Pelayanan cepat belum ada hari sudah selesai Semoga tetap bisa memberikan pelayanan yg terbaik setiap harinya dengan sepenuh hati

Selanjutnya dilakukan perubahan huruf besar menjadi huruf kecil (*case folding*) agar memudahkan dalam identifikasi kata. Hasil *Caselolding* dapat terlihat pada tabel 10.

**Tabel 10.** Hasil *Case Folding*

Sebelum	Hasil Pembersihan Data
Pelayanan cepat belum ada hari sudah selesai Semoga tetap bisa memberikan pelayanan yg terbaik setiap harinya dengan sepenuh hati	pelayanan cepat belum ada hari sudah selesai semoga tetap bisa memberikan pelayanan yg terbaik setiap harinya dengan sepenuh hati

## 2. Normalisasi kata

Tahapan ini mengubah kata tidak baku menjadi kata baku dengan dataset panduan. Pada penelitian ini memanfaatkan dataset kamusbaku. Hasil normalisasi ditunjukkan pada tabel 11.

**Tabel 11.** Hasil Normalisasi

Data Sebelum Normalisasi	Hasil Normalisasi
pelayanan cepat belum ada hari sudah selesai semoga tetap bisa memberikan pelayanan yg terbaik setiap harinya dengan sepenuh hati	pelayanan cepat belum ada hari sudah selesai semoga tetap bisa memberikan pelayanan yang terbaik setiap harinya dengan sepenuh hati

### 3. *Tokenize*

*Tokenize* adalah proses memecah data berupa teks dari sebuah kalimat menjadi bagian berupa potongan-potongan kata.

**Tabel 12. Hasil Tokenize**

Data Sebelum Tokenize	Hasil Tokenize
pelayanan cepat belum ada hari sudah selesai semoga tetap bisa memberikan pelayanan yang terbaik setiap harinya dengan sepenuh hati	['pelayanan', 'cepat', 'belum', 'ada', 'hari', 'sudah', 'selesai', 'semoga', 'tetap', 'bisa', 'memberikan', 'pelayanan', 'yang', 'terbaik', 'setiap', 'harinya', 'dengan', 'sepenuh', 'hati']

#### 4. Stopwords

Stopwords merupakan proses menghapus kata-kata yang kurang/tidak memiliki arti/makna untuk mempengaruhi sentimen dan bukan merupakan deskripsi sesuatu yang semestinya dihilangkan. Proses *stopword* yang digunakan pada penelitian ini memanfaatkan *library Sastrawi* yang di dalamnya terdapat corpus *stopword* bahasa Indonesia dan menambah secara manual beberapa kata yang belum ada dalam *library Sastrawi*.

**Tabel 13.** Hasil Stopwords

Data Sebelum Stopword	Hasil Stopword
['pelayanan', 'cepat', 'belum', 'ada', 'hari', 'sudah', 'selesai', 'semoga', 'tetap', 'bisa', 'memberikan', 'pelayanan', 'yang', 'terbaik', 'setiap', 'harinya', 'dengan', 'sepenuh', 'hati']	['pelayanan', 'cepat', 'hari', 'selesai', 'semoga', 'tetap', 'memberikan', 'pelayanan', 'terbaik', 'harinya', 'sepenuh', 'hati']

## 5. Stemming

Pada tahap ini, kata-kata dalam komentar diubah ke bentuk dasarnya (stem) dengan menghapus imbuhan atau akhiran. Langkah ini bertujuan untuk mengurangi variasi kata yang memiliki akar kata yang sama.

**Tabel 14.** Hasil *Stemming*

Data sebelum <i>Stemming</i>	<i>Hasil Stemming</i>
['pelayanan', 'cepat', 'hari', 'selesai', 'semoga', 'tetap', 'memberikan', 'pelayanan', 'terbaik', 'harinya', 'sepenuh', 'hati']	layan cepat hari selesai moga tetap beri layan baik hari sepenuh hati

b) Pelabelan

Pada penelitian ini dilakukan pelabelan menggunakan dua metode yang berbeda yaitu *Inset Lexicon* dan *IndoBERT*. *InSet Lexicon* atau *Indonesian Sentiment Lexicon* adalah sebuah kamus leksikal khusus untuk analisis sentimen dalam bahasa Indonesia yang berisi daftar kata-kata beserta bobot nilai polaritas sentimen masing-masing kata. *IndoBERT* adalah model pretrained transformer berbasis arsitektur *BERT* yang telah dilatih khusus untuk bahasa Indonesia yang digunakan untuk ekstraksi fitur teks dan pelabelan otomatis teks. Dalam penelitian ini data diklasifikasikan menjadi sentimen positif dan negatif.

### 1) Pelabelan dengan *InSet Lexicon*

Pada penelitian ini digunakan metode pelabelan *InSet Lexicon* atau *Indonesian Sentiment Lexicon* sebagai kamus dalam proses pelabelan. Kamus *InSet Lexicon* terdiri dari 3.609 kata positif dan 6.609 kata negatif dengan bobot tantara -5 sampai +5. Kamus *InSet Lexicon* disimpan dalam berkas kamus\_positive.xlsx memuat daftar kata yang dinilai memiliki makna positif berdasarkan *InSet Lexicon*, sedangkan kamus\_negative.xlsx berisi daftar kata yang memiliki makna negatif. Kedua daftar ini kemudian dikonversi menjadi set (kumpulan elemen unik) dan seluruh katanya diubah menjadi huruf kecil, sehingga proses pencocokan kata tidak membedakan antara huruf besar dan huruf kecil.

Label sentimen ditentukan berdasarkan total nilai *polarity*. Jika nilai di atas 0, maka saran tersebut termasuk sentimen positif. Jika nilai total di bawah 0, maka saran termasuk sentimen negatif. Sedangkan jika total nilai polarity adalah 0, maka saran termasuk sentimen netral. Dalam penelitian ini pelabelan disederhanakan menjadi 2 kelas positif dan negatif, tidak menggunakan label netral dengan pertimbangan bahwa pada sentimen netral biasanya dianggap tidak memberikan informasi yang cukup untuk analisis emosi atau sikap dan sering kali tumpang tindih dengan positif atau negatif dalam hal kosakata.

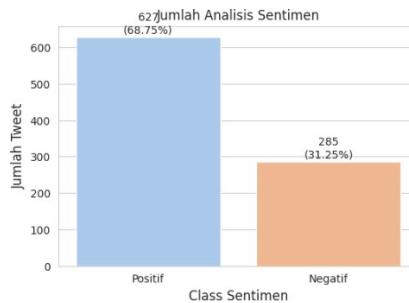
```

1 import pandas as pd
2 import requests
3 import random
4
5 positive_lexicon = set(pd.read_excel("kamus_positive.xlsx", header=None)
6 [0].str.lower())
6 negative_lexicon = set(pd.read_excel("kamus_negative.xlsx", header=None)
7 [0].str.lower())
7
8 #Fungsi untuk menentukan sentimen dan menghitung skornya
9 def determine_sentiment(text):
10     if isinstance(text, str):
11         positive_count = sum(1 for word in text.split() if word in
12         positive_lexicon)
13         negative_count = sum(1 for word in text.split() if word in
14         negative_lexicon)
15         sentimen_score = positive_count - negative_count
16
17         if sentimen_score >= 0:
18             sentiment = "Positif"
19         else:
20             sentiment = "Negatif"
21
22         return sentimen_score, sentiment
23     return 0, "Positif"
24
25 #tentukan sentimen dan skor untuk setiap ulasan
26 data[['Score', 'Sentimen']] = data['stemmed'].apply(lambda x:pd.Series
27 (determine_sentiment(x)))
28
29 data.head(20)

```

Gambar 4. Kode Program Pelabelan *InSet Lexicon*

Grafik hasil pelabelan *InSet Lexicon* dengan membagi ke dalam 2 (dua) kelas positif dan negatif dapat dilihat pada Gambar 5.



Gambar 5. Grafik Hasil Pelabelan *InSet Lexicon*

Pelabelan *InSet Lexicon* menghasilkan jumlah sentimen 627 di kelas positif dan 285 di kelas negatif.

## 2) Pelabelan dengan IndoBERT

Pada tahap ini dilakukan pelabelan pelabelan *IndoBERT* dari *w11wo/indonesian-roberta-base-sentiment-classifier* yang merupakan model *IndoBERT* yang sudah dilatih (*fine-tuned*) menggunakan data sentimen bahasa Indonesia. Seperti halnya pelabelan dengan *Inset Lexicon* yang dibagi menjadi 2 (dua) kelas, pada pelabelan menggunakan *IndoBERT*, data sentimen diklasifikasikan menjadi 2 (dua) kelas positif dan negatif.

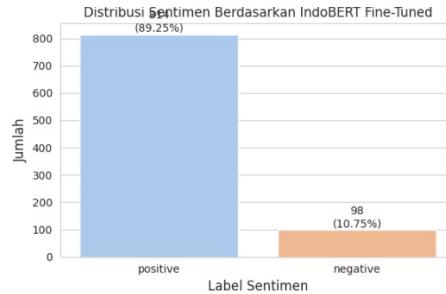
```

1 from transformers import AutoTokenizer, AutoModelForSequenceClassification,
2 pipeline
3
4 # Load model IndoBERT untuk sentimen
5 model_name = "w11wo/indonesian-roberta-base-sentiment-classifier"
6 tokenizer = AutoTokenizer.from_pretrained(model_name)
7 model = AutoModelForSequenceClassification.from_pretrained(model_name)
8
9 # Buat pipeline sentimen
10 sentiment_analysis = pipeline("sentiment-analysis", model=model,
11                             tokenizer=tokenizer)
12
13 # Fungsi pelabelan dengan IndoBERT
14 def label_with_sentiment_model(text):
15     try:
16         result = sentiment_analysis(text)[0]
17         return result['label'] # sudah dalam 'positive', 'neutral', atau
18         'negative'
19     except:
20         return 'neutral' # fallback jika error
21
22 # Terapkan pelabelan sentimen
23 data['indobert_label'] = data['stemmed'].apply(label_with_sentiment_model)
24
25 # Mapping: neutral dianggap positif
26 data['indobert_label'] = data['indobert_label'].replace({'neutral': 'positive'})
27
28 # Cek distribusi label setelah mapping
29 print("Distribusi label setelah mapping neutral + positive:")
30 print(data['indobert_label'].value_counts())
31
32 # (Opsiional) Simpan ke file CSV
33 data.to_csv("hasil_label_indobert.csv", index=False)
34

```

Gambar 6. Kode Program untuk Pelabelan *IndoBERT*

Hasil pelabelan menunjukkan dari 921 dataset dihasilkan 814 data berlabel positif dan 98 data berlabel negatif. Hasil Pelabelan *IndoBERT* dengan bentuk bagan dapat dilihat pada gambar 7.

Gambar 7. Grafik Sebaran Kelas Pelabelan *IndoBERT*

## c) Pembobotan

Pembobotan dilakukan menggunakan metode TF-IDF, dimana TF(*Term Frequency*) menghitung frekuensi kemunculan satu kata dalam dokumen. Sementara itu, IDF (*Inverse Document Frequency*) bertujuan untuk menentukan pentingnya sebuah kata dalam dokumen berdasarkan kategorinya, sekaligus mengidentifikasi kata-kata yang serupa dengan kategori tersebut. Hasil perhitungan *Prior Probabilitas* kelas positif dan negatif pada pelabelan *InSet Lexicon* menghasilkan Probabilitas sentimen positif dan negatif sebagaimana Tabel 15.

Tabel 15. Hasil *Stemming*

Sentimen	Probabilitas
Positif	0,6875
Negatif	0,3125

Hasil *Probabilitas Likelihood* ditunjukkan pada pelabelan *InSet Lexicon* ditunjukkan pada tabel 16.

Tabel 16. Hasil *Probabilitas Likelihood*

Term	P(wi positif)	P(wi negatif)
Tracking	0,0021	0,0008
Perizinan	0,0082	0,0123
Mudah	0,0313	0,0145
Notifikasi	0,0027	0,0008
Rinci	0,0021	0,0008
Pelayanan	0,4112	0,0096
Lama	0,0031	0,0058
Waktu	0,0077	0,0031
Estimasi	0,0047	0,0021
Baik	0,0427	0,0436

Perhitungan Hasil *posterior probabilitas* dokumen yang mengandung *term* perizinan, mudah dan waktu sebagai berikut :

$$\begin{aligned}
 P(\text{Positif}|\text{Dokumen}) &= P(\text{Positif}) \times P(\text{perizinan}|\text{Positif}) \times P(\text{mudah}|\text{positif}) \times P(\text{waktu}|\text{positif}) \\
 &= 0,667 \times 0,0082 \times 0,313 \times 0,0077 \\
 &= 0,00001318
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Negatif}|\text{Dokumen}) &= P(\text{Negatif}) \times P(\text{perizinan}|\text{Negatif}) \times P(\text{mudah}|\text{Negatif}) \times P(\text{waktu}|\text{Negatif}) \\
 &= 0,333 \times 0,0123 \times 0,0145 \times 0,0031 \\
 &= 0,0000001841
 \end{aligned}$$

Hasil prediksi dokumen dapat dilihat pada tabel 17.

Tabel 17. Hasil Prediksi Dokumen

P(Positive Dokumen)	P(Negative Dokumen)
0,00001318	0,0000001841

Hasil prediksi dokumen dengan isi *term* [perizinan, mudah, waktu] menunjukkan angka lebih tinggi pada kelas positif disbanding kelas negatif, sehingga menghasilkan prediksi positif.

## d) Pengujian Klasifikasi



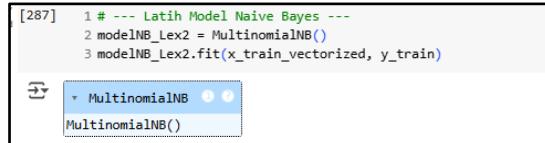
Pengujian dilakukan terhadap model klasifikasi yang digunakan. Pada penelitian ini, digunakan metode *Naïve Bayes* sebagai pendekatan analisis. *Naïve Bayes* merupakan salah satu teknik klasifikasi yang sangat dikenal dalam bidang machine learning dan data mining karena efektivitasnya. Langkah yang dilakukan yaitu :

1. Split Data (Membagi Data Latih dan Data Uji)

Pada penelitian ini data dibagi menjadi data latih dan data uji dengan rasio 90:10. Data latih dan data uji pada data “stemmed” diberi nama *x\_train* dan *x\_tes* sedangkan pada data “Sentimen” diberi nama *y\_train* dan *y\_test*. Dari dataset sejumlah 921 data terbagi menjadi 820 data latih dan 92 data uji.

2. Permodelan *Naïve Bayes*

Setelah dilakukan pembagian data latih dan data uji, data akan dilakukan untuk melatih model klasifikasi menggunakan *MultinomialNB()* dari library *sklearn.naive\_bayes*. Kemudian akan dilakukan prediksi pada data uji hasil vektorisasi *TF\_IDF*.



```
[287] 1 # --- Latih Model Naive Bayes ---
2 modelNB_Lex2 = MultinomialNB()
3 modelNB_Lex2.fit(x_train_vectorized, y_train)
```

**Gambar 8.** Kode Program Model *Naïve Bayes*

3. Evaluasi Model

Pada tahap ini dilakukan evaluasi model dengan untuk mengetahui akurasi model *Naïve Bayes* serta nilai *precision*, *recall*, *f1-score* pada masing-masing kelas. Hasil akurasi model *Naïve Bayes* dengan pelabelan *InSet Lexicon* dan *IndoBERT* dapat dilihat pada gambar 9 dan gambar 10.

Akurasi Naive Bayes: 89.13%				
Naive Bayes Classification Report:				
	precision	recall	f1-score	support
Negatif	0.95	0.69	0.80	29
Positif	0.87	0.98	0.93	63
accuracy			0.89	92
macro avg	0.91	0.84	0.86	92
weighted avg	0.90	0.89	0.89	92

**Gambar 9.** Hasil Evaluasi Model *Naïve Bayes* dengan Pelabelan *InSet Lexicon*

Hasil evaluasi model *Naïve Bayes* dengan pelabelan *InSet Lexicon* menunjukkan nilai akurasi 89,13. Nilai *precision* sebesar 0,95, *recall* sebesar 0,69, *f1-score* sebesar 0,80 pada kelas negatif dan pada kelas positif nilai *precision* menunjukkan angka sebesar 0,87, *recall* sebesar 0,98 dan *f1-score* sebesar 0,93.

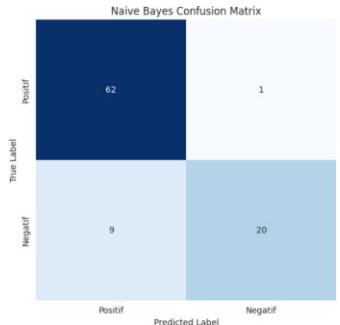
Akurasi Naive Bayes: 90.22%				
Naive Bayes Classification Report:				
	precision	recall	f1-score	support
negative	1.00	0.10	0.18	10
positive	0.90	1.00	0.95	82
accuracy			0.90	92
macro avg	0.95	0.55	0.56	92
weighted avg	0.91	0.90	0.86	92

**Gambar 10.** Hasil Evaluasi Model *Naïve Bayes* dengan Pelabelan *IndoBERT*

Hasil evaluasi model *Naïve Bayes* dengan pelabelan *IndoBERT* menunjukkan nilai akurasi 90,22. Nilai *precision* sebesar 1,00, *recall* sebesar 0,10, *f1-score* sebesar 0,18 pada kelas negatif dan pada kelas positif nilai *precision* menunjukkan angka sebesar 0,90, *recall* sebesar 0,10 dan *f1-score* sebesar 0,95.

4. Confusion Matrix

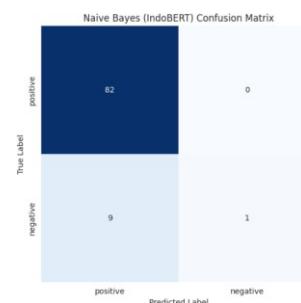
Pada tahap ini dilakukan pengujian model *Naïve Bayes* menggunakan *Confusion Matrix*. *Confusion matrix* menyajikan performa klasifikasi model melalui tabulasi prediksi benar dan salah per kelas. Proses evaluasi diawali dengan pembagian dataset menjadi *training* dan *testing set*. Model yang telah dilatih kemudian diuji menggunakan data *testing*, dimana hasil prediksinya dibandingkan dengan label sebenarnya untuk menghasilkan matriks *True Positif*(TP), *True Negatif*(TN), *False Positif*(FP), dan *False Negatif*(FN). Hasil pengujian dengan *Confusion Matrix* pada model *Naïve Bayes* dengan pelabelan *InSet Lexicon* dapat dilihat pada gambar 11.



**Gambar 11.** Hasil Confusion Matrix Model *Naive Bayes* dengan Pelabelan *InSet Lexicon*

Berdasarkan hasil pengujian, diperoleh bahwa pada *True Positive (TP)* terdapat 62 data positif berhasil diprediksi dengan benar sebagai positif dan pada *False Negative (FN)* terdapat 1 data positif diprediksi salah sebagai negatif. Sedangkan *True Negative (TN)* terdapat 20 data negatif berhasil diprediksi dengan benar sebagai negatif dan pada *False Positive (FP)* terdapat 9 data negatif diprediksi salah sebagai positif.

Hasil pengujian dengan *Confussion Matrix* pada model *Naïve Bayes* dengan pelabelan *IndoBERT* dapat dilihat pada gambar 12.



**Gambar 12.** Hasil Confusion Matrix Model Naive Bayes dengan Pelabelan IndoBERT

Berdasarkan hasil pengujian, diperoleh bahwa pada *True Positive (TP)* terdapat 82 data positif berhasil diprediksi dengan benar sebagai positif dan pada *False Negative (FN)* bernilai 0 artinya tidak ada data yang positif diprediksi salah sebagai negatif. Sedangkan *True Negative (TN)* terdapat 1 data negatif berhasil diprediksi dengan benar sebagai negatif dan pada *False Positive (FP)* terdapat 9 data negatif diprediksi salah sebagai positif.

### 3.2 Pembahasan

#### 4.2.1 Hasil Pelabelan

Penerapan metode pelabelan *InSet Lexicon* dan *IndoBERT* dalam menganalisis sentimen pelayanan publik pada DPMPTSP Kabupaten Magelang pada data saran terpilih yang berjumlah 921 data mendapatkan hasil bahwa pelabelan *InSet Lexicon* menghasilkan jumlah 627 sentimen di kelas positif dan 285 di kelas negatif. Sedangkan pelabelan menggunakan *IndoBERT* menghasilkan 814 kelas positif 98 kelas negatif. Hal ini menunjukkan bahwa tren saran pada Survei Kepuasan Masyarakat didominasi dengan sentimen positif dibandingkan dengan sentimen negatif.



**Gambar 13.** Hasil *WorldCloud* Kelas Positif *InSet Lexicon*

Terdapat **kata-kata dominan** seperti “layan”, “baik”, “bagus”, “bantu”, “mudah”, “selesai”, “cepat”, memberikan makna masyarakat banyak memberikan apresiasi terhadap pelayanan secara umum bahwa kualitas pelayanan memuaskan. Berdasarkan hasil analisis sentimen, terdapat tiga unsur utama yang paling dominan dalam membentuk sentimen positif. **Pertama, unsur waktu penyelesaian** yang ditunjukkan melalui kemunculan kata-kata

seperti “cepat”, “selesai” dan “tepat waktu”, mengindikasikan kepuasan pengguna terhadap ketepatan dan kecepatan layanan. **Kedua, unsur sistem dan prosedur** yang tercermin dari kata “mudah”, “online”, “proses”, “sidering” menggambarkan persepsi positif terhadap kemudahan akses dan kelancaran proses pelayanan. **Ketiga, unsur perilaku petugas** yang terlihat dari kata “baik”, “bantu”, dan “ramah”, menunjukkan apresiasi masyarakat terhadap sikap, keramahan, serta bantuan yang diberikan oleh petugas.



**Gambar 14.** Hasil WorldCloud Kelas Negatif InSet Lexicon

Kata-kata yang dominan pada label negatif yaitu “cukup”, “bagus”, “baik”. Walaupun terkesan positif, namun kata “cukup” disini bermakna lumayan tetapi belum memuaskan, masyarakat mungkin merasa pelayanan belum optimal meskipun tidak buruk. Terdapat kata “keluh”, ditindaklanjuti, “mohon” merupakan sering muncul, menunjukkan masalah respon terhadap keluhan.

#### 4.2.2 Klasifikasi Model *Naïve Bayes* dengan Label *InSet Lexicon* dan *IndoBERT*

Klasifikasi sentimen pelayanan publik berdasarkan saran pada Survei Kepuasan Masyarakat Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Kabupaten Magelang menggunakan *Naïve Bayes* dengan pendekatan pelabelan *InSet Lexicon* dan *IndoBERT* mendapatkan hasil akurasi model *Naïve Bayes* dengan label *InSet Lexicon* dan *IndoBERT* yang ditunjukkan pada tabel 18.

**Tabel 18.** Perbandingan Akurasi Model

Model Pelabelan <i>Naïve Bayes</i>	Kelas	Akurasi	Presisi	<i>Recall</i>	<i>F-1 Score</i>
<i>InSet Lexicon</i>	Positif	89,13	0.87	0.98	0.93
	Negatif		0.95	0.69	0.80
<i>IndoBERT</i>	Positif	90,22	0.90	1.00	0.95
	Negatif		1.00	0.10	0.18

Dari tabel diatas dapat dilihat bahwa model *Naïve Bayes* dengan pelabelan *InSet Lexicon* mendapatkan akurasi 89,13 sedangkan *IndoBERT* berada diatas *InSet Lexicon* pada angka 90,22. *IndoBERT* unggul pada kelas positif dengan nilai presisi 0.9, recall 1.00 dan *f1-score* 0.95 tetapi lemah di kelas negatif dengan angka presisi 1.00, recall 0.10 dan *f1-score* 0.18. Sedangkan *InSet Lexicon* memiliki akurasi yang lebih seimbang pada kedua kelas. Pada kelas positif nilai presisi berada di angka 0.87, recall 0.98 dan *f1-score* 0.93 sedangkan kelas negatif nilai presisi berada pada angka 0.95, recall 0.69 dan *f1-score* 0.80. Dengan demikian dapat dikatakan bahwa penggunaan pelabelan *InSet Lexicon* dan *IndoBERT* dapat berpengaruh terhadap akurasi dari model analisis sentimen *Naïve Bayes*.

#### 4. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat diambil kesimpulan bahwa penerapan metode pelabelan *InSet Lexicon* dan *IndoBERT* dalam menganalisis sentimen pelayanan publik pada DPMPTSP Kabupaten Magelang menggunakan *Naïve Bayes* mendapatkan hasil bahwa saran pada SKM di DPMPTSP Kabupaten Magelang didominasi oleh sentimen positif. Terdapat tiga unsur utama yang paling dominan dalam membentuk sentimen positif yaitu pertama, unsur waktu penyelesaian, unsur sistem dan prosedur, dan unsur perilaku pelaksana. Hasil akurasi model klasifikasi *Naïve Bayes* dengan pelabelan *InSet Lexicon* menunjukkan nilai 89,13 sedangkan model *Naïve Bayes* dengan pelabelan *IndoBERT* menunjukkan nilai akurasi 90,22. Hal ini menunjukkan bahwa model *Naïve Bayes* efektif digunakan sebagai metode dalam melakukan analisis sentimen pada pelayanan publik.

Penulis menyampaikan saran yang diharapkan menjadi masukan pengembangan penelitian selanjutnya yaitu pertama disarankan penelitian berikutnya dapat menggunakan komentar masyarakat yang berasal dari media sosial. Kedua, terdapat ketidakseimbangan jumlah kelas positif dan kelas negatif serta adanya distribusi data latih yang tidak seimbang, diharapkan pada penelitian selanjutnya dapat memanfaatkan oversampling kelas minoritas, misalnya dengan SMOTE (*Synthetic Minority Oversampling Technique*). Yang terakhir, walaupun metode *Naïve Bayes* dalam penelitian

ini digunakan karena kesederhanan dan efisiennya, akan lebih baik jika pada penelitian selanjutnya dilakukan perbandingan dengan algoritma *machine learning* lain untuk melihat efektifitas terhadap jenis data yang digunakan.

## **UCAPAN TERIMAKASIH**

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

## **REFERENCES**

- [1] *UU Nomor 25 Tahun 2009*. 2009.
- [2] P. R. Indonesia, *Peraturan Pemerintah Republik Indonesia Nomor 96 Tahun 2021*. 2012. [Online]. Available: [www.djpp.depumham.go.id](http://www.djpp.depumham.go.id)
- [3] P. D. K. Perbup Magelang No.27, *Peraturan Bupati Magelang Nomor 27 Tahun 2025*, vol. 1, no. 22. 2025, pp. 24–25.
- [4] P. B. M. N. 19, *Peraturan Bupati Magelang No. 19 Tahun 2023*. 2023, pp. 1–58.
- [5] Permenpan No.14, *Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia Nomor 14*, vol. 94, no. 2. 2017, pp. 459–464. doi: 10.1016/0014-4827(75)90518-2.
- [6] A. Firdaus, W. I. Firdaus, P. Studi, T. Informatika, M. Digital, and P. N. Sriwijaya, “Text Mining dan Pola Algoritma Dalam Penyelesaian Masalah Informasi (Sebuah Ulasan),” *JUPITER*, vol. 13, no. 1, pp. 66–78, 2021.
- [7] P. Y. Saputra, D. H. Subhi, F. Zain, and A. Winatama, “Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes,” *J. Inform. Polinema*, 2023.
- [8] S. Indriani and M. Syafrullah, “Multinomial Naïve Bayes Untuk Menganalisis Sentimen Layanan Jasa Ekspedisi Sicepat Ekspres,” *Semin. Nas. Mhs. Fak. Teknol. Inf. Jakarta-Indonesia*, no. September, pp. 445–453, 2022, [Online]. Available: <https://developer.twitter.com/>
- [9] R. R. Putri and N. Cahyono, “Analisis Sentimen Komentar Masyarakat Terhadap Pelayanan Publik Pemerintah Dki Jakarta Dengan Algoritma Super Vector Machine Dan Naive Bayes,” *J. Mhs. Tek. Inform.*, vol. 8, no. 2, p. 2363, 2024.
- [10] C. Rahmawati and P. Sukmasetya, “Sentimen Analisis Opini Masyarakat Terhadap Kebijakan Kominfo atas Pemblokiran Situs non-PSE pada Media Sosial Twitter,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1393, Oct. 2022, doi: 10.30865/jurikom.v9i5.4950.
- [11] D. Darwis, E. Shintya Pratiwi, A. Ferico, and O. Pasaribu, “Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *J. Ilm. Edutic*, vol. 7, no. 1, 2020.
- [12] I. Gede Bintang Arya Budaya, L. Putu Safitri Pratiwi, and D. Panji Agustino, “Klasifikasi Sentimen untuk Analisis Kepuasan Pelayanan Puskesmas Berbasis Arsitektur LSTM,” *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 12, no. 4, 2023, doi: 10.30591/smartcomp.v12i4.5361.
- [13] A. Khusnul Khotimah, “Analisis Sentimen Terhadap Kualitas Pelayanan,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 3044–3048, 2024, doi: 10.36040/jati.v8i3.9520.
- [14] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [15] D. Rika Widianita, “Analisis Sentimen Aplikasi Reksadana Berdasarkan Ulasan Di Play Store Menggunakan Metode Naive Bayes,” 2023.
- [16] B. R. Tsania Dzulkarnain, Dian Eka Ratnawati, “Penggunaan Metode Naïve Bayes Classifier Pada Analisis the Use of the Naïve Bayes Classifier Method in Sentiment Analysis of the Community ’ S Assessment of Hospital Services in,” vol. 10, no. 7, pp. 1453–1460, 2023, doi: 10.25126/jtiik.2024117979.
- [17] H. Firda *et al.*, “Sistemasi: Jurnal Sistem Informasi Perbandingan Pelabelan Rating-based dan Inset Lexicon-based dalam Analisis Sentimen Menggunakan SVM (Studi Kasus: Ulasan Aplikasi GoBiz di Google Play Store) Comparison of Rating-based and Inset Lexicon-based Labeling in,” *Sist. J. Sist. Inf.*, vol. 14, pp. 516–528, 2025, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [18] S. N. Erma Susanti, Maimunah, “Sentiment Snalysis of youtube comments Using Machine Learning Models,” *Penelit. Ilmu Komputer, Sist. Embed. Log.*, vol. 13, no. 225, pp. 1011–1014, 2025, doi: 10.33558/piksel.v13i1.10743.
- [19] R. Wati and S. Ernawati, “Analisis Sentimen Persepsi Publik Mengenai PPKM Pada Twitter Berbasis SVM Menggunakan Python,” 2021. [Online]. Available: <https://netlytic.org>
- [20] M. Namira, “Analisis Sentimen Vaksinasi Covid-19 Pada Komentar Youtube Degan Menggunakan Algoritma Naive Bayes Classifier (NBC) Dan Support Vector Machine (SVM),” 2023.

