



## **SURAT TUGAS**

Nomor : 765/D/PK/2026

***Assalamu'alaikum warahmatullahi wabarakatuh,***

Pimpinan Fakultas Teknologi Industri dan Informatika Universitas Muhammadiyah  
Prof. DR. HAMKA, memberikan tugas kepada:

- Nama : **1. Firman Noor Hasan, S.Kom., M.TI**  
**2. Ir. Ade Davy Wiranata, S.Kom., M.Kom**  
**3. Ir. Dimas Febriawan, S.Kom., M.TI**  
**4. Mia Kamayani, ST., MT**  
**5. Estu Sinduningrum, ST., MT**
- Tugas : Sebagai Dosen Penguji Sidang Skripsi Program Studi Teknik Informatika.
- Waktu : Jum'at, 24 April 2026
- Tempat : R.304 (Ruang Smart Class)
- Lain-lain : Setelah melaksanakan tugas agar memberikan laporan secara tertulis kepada Pimpinan Fakultas Teknologi Industri dan Informatika UHAMKA.

Demikian surat tugas ini disampaikan, agar dapat dilaksanakan dengan sebaik-baiknya sebagai amanah dan ibadah kepada Allah SWT.

***Wabillahit taufiq walhidayah,***  
***Wassalamu'alaikum warahmatullahi wabarakatuh***

Jakarta, 22 April 2026 M  
05 Zulkaidah 1447 H



Dekan,

**Dr. Ir. Dan Mugisidi, ST., M.Si.**

**Tembusan :**

1. Wakil Dekan I & II;
2. KTU u.p. Kasubbang Keuangan;  
FT UHAMKA.



**Lampiran Surat Tugas Dekan FTII UHAMKA**

Nomor : 765/D/PK/2026

Tanggal : 22 April 2026 M/05 Zulkaidah 1447 H

**DAFTAR NAMA PENGUJI DAN PESERTA SIDANG SKRIPSI  
PROGRAM STUDI TEKNIK INFORMATIKA FTII UHAMKA**

| No | Nama                   | NIM        | Judul                                                                                                                                                                         | Pembimbing                                | Penguji 1                             | Penguji 2                                  | Hari,<br>Tanggal         | Jam         |
|----|------------------------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|---------------------------------------|--------------------------------------------|--------------------------|-------------|
| 1  | Fakhri Huda<br>Awalin  | 1903015034 | ANALISIS KERENTANAN<br>KEAMANAN WEBSITE<br>PT. TIGER WONG<br>ENTERTAINMENT<br>DENGAN METODE<br>BLACK-BOX<br>MENGUNAKAN OWASP<br>ZAP DAN WPSCAN                                | Dimas<br>Febriawan,<br>S.Kom., M.T.I.     | Firman Noor<br>Hasan, S.Kom.,<br>M.TI | Ade Davy<br>Wiranata,<br>S.Kom.,<br>M.Kom. | Jum'at, 24<br>April 2026 | 08.30-09.30 |
| 2  | Yogga Tolly<br>Dewanto | 2003015136 | ANALISIS SENTIMEN<br>TERHADAP MINAT<br>MASYARAKAT JAKARTA<br>YANG MEMILIH<br>KENDARAAN UMUM<br>MENGUNAKAN<br>ALGORITMA NAÏVE<br>BAYES                                         | Ade Davy<br>Wiranata,<br>S.Kom., M.Kom.   | Firman Noor<br>Hasan, S.Kom.,<br>M.TI | Mia Kamayani<br>S.T., M.T.                 | Jum'at, 24<br>April 2026 | 09.45-10.45 |
| 3  | Hilmi Ammar            | 2003015162 | PERBANDINGAN<br>TINGKAT AKURASI<br>ALGORITMA NAIVE<br>BAYES, DECISION TREE<br>DAN RANDOM FOREST<br>DALAM ANALISIS<br>SENTIMEN PENGGUNA<br>APLIKASI SAMSAT<br>DIGITAL NASIONAL | Ade Davy<br>Wiranata,<br>S.Kom., M.Kom.   | Firman Noor<br>Hasan, S.Kom.,<br>M.TI | Estu<br>Sinduningrum,<br>S.T., M.T.        | Jum'at, 24<br>April 2026 | 11.00-12.00 |
| 4  | Putri Salma            | 2203015097 | SISTEM E-COMMERCE<br>DENGAN INTEGRASI<br>FORECASTING<br>PENJUALAN<br>MENGUNAKAN<br>METODE PERSONAL<br>EXTREME<br>PROGRAMMING (STUDI<br>KASUS: JKT SWEET<br>DESSERT)           | Atiqah Meutia<br>Hilda, S.Kom.,<br>M.Kom. | Mia Kamayani<br>S.T., M.T.            | Estu<br>Sinduningrum,<br>S.T., M.T.        | Jum'at, 24<br>April 2026 | 13.00-14.00 |



Dekan

**Dr. Ir. Dan Mugisidi, ST., M.Si.**

**JADWAL SIDANG SKRIPSI TEKNIK INFORMATIKA  
SEMESTER GENAP 2025/2026 (BATCH 1-APRIL 2026)**

| No | Nama Lengkap        | NIM        | Jenis Tugas Akhir | Judul Skripsi/Non-Skripsi                                                                                                                             | Dosen Pembimbing                    | Penguji 1                       | Penguji 2                         | Hari, tanggal         | Pukul       | Room     |
|----|---------------------|------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------|---------------------------------|-----------------------------------|-----------------------|-------------|----------|
| 1  | Fakhri Huda Awaln   | 1903015034 | Skripsi           | Analisis Kerentanan Keamanan Website PT. Tiger Wong Entertainment dengan Metode Black-Box Menggunakan OWASP ZAP dan WPSCAN                            | Dimas Febriawan, S.Kom., M.T.I.     | Firman Noor Hasan, S.Kom., M.TI | Ade Davy Wiranata, S.Kom., M.Kom. | Jum'at, 24 April 2026 | 08.30-09.30 | R.FT 304 |
| 2  | Yogga Tolly Dewanto | 2003015136 | Artikel Jurnal    | ANALISIS SENTIMEN TERHADAP MINAT MASYARAKAT JAKARTA YANG MEMILIH KENDARAAN UMUM MENGGUNAKAN ALGORITMA NAÏVE BAYES                                     | Ade Davy Wiranata, S.Kom., M.Kom.   | Firman Noor Hasan, S.Kom., M.TI | Mia Kamayani S.T., M.T.           | Jum'at, 24 April 2026 | 09.45-10.45 | R.FT 304 |
| 3  | Hilmi Ammar         | 2003015162 | Artikel Jurnal    | Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional | Ade Davy Wiranata, S.Kom., M.Kom.   | Firman Noor Hasan, S.Kom., M.TI | Estu Sinduningrum, S.T., M.T.     | Jum'at, 24 April 2026 | 11.00-12.00 | R.FT 304 |
| 4  | Putri Salma         | 2203015097 | HAKI              | SISTEM E-COMMERCE DENGAN INTEGRASI FORECASTING PENJUALAN MENGGUNAKAN METODE PERSONAL EXTREME PROGRAMMING (STUDI KASUS: JKT SWEET DESSERT)             | Atiqah Meutia Hilda, S.Kom., M.Kom. | Mia Kamayani S.T., M.T.         | Estu Sinduningrum, S.T., M.T.     | Jum'at, 24 April 2026 | 13.00-14.00 | R.FT 304 |

Jakarta, 21 April 2026  
Kaprod Teknik Informatika

  
Mia Kamayani, S.T., M.T.

|                                                                                         |                                |                                             |
|-----------------------------------------------------------------------------------------|--------------------------------|---------------------------------------------|
|        | <b>LEMBAR<br/>BERITA ACARA</b> | Form No :                                   |
|                                                                                         |                                | 15/11/Prodi-TI/Akad/2025                    |
|                                                                                         |                                | <b>Program Studi<br/>Teknik Informatika</b> |
| Fakultas Teknologi Industri Dan Informatika<br>Universitas Muhammadiyah Prof. DR. HAMKA |                                |                                             |

*Bismillahirrahmaanirrahiim*

Pada hari ini Jumat, 24 April 2026 telah diadakan ujian Tugas Akhir / Skripsi Mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Industri Dan Informatika UHAMKA

Dengan Susunan sebagai berikut :

|   |                                       |                             |
|---|---------------------------------------|-----------------------------|
| 1 | Firman Noor Hasan, S.Kom., M.TI.      | Ketua Sidang/ Penguji 1     |
| 2 | Estu Sinduningrum, S.T., M.T.         | Anggota Sidang/penguji 2    |
| 3 | Ir. Ade Davy Wiranata, S.Kom., M.Kom. | Anggota Sidang/Pembimbing 1 |
| 4 |                                       | Anggota Sidang/Pembimbing 2 |

Dengan peserta ujian :

|               |                    |             |                   |
|---------------|--------------------|-------------|-------------------|
| <b>Nama :</b> | <b>Hilmi Ammar</b> | <b>NIM:</b> | <b>2003015162</b> |
|---------------|--------------------|-------------|-------------------|

|                       |                                                                                                                                                              |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Judul Skripsi:</b> | <b>Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional</b> |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|

Nilai ujian Penguji & Pembimbing

|              |              |               |
|--------------|--------------|---------------|
| 1            | Penguji 1    | <b>79</b>     |
| 2            | Penguji 2    | <b>68.5</b>   |
| 3            | Pembimbing 1 | <b>84.5</b>   |
| 4            | Pembimbing 2 |               |
| <b>Nilai</b> |              | <b>79.125</b> |

|                                    |              |
|------------------------------------|--------------|
| Peserta sidang tersebut dinyatakan | <b>LULUS</b> |
| Dengan Predikat Nilai              | <b>B</b>     |

Demikian berita acara ini dibuat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

Ketua Program Studi

  
**Mia Kamayani Sulaeman, S.T., M.T.**

Jumat, 24 April 2026

Panitia Ujian TA / Skripsi

Ketua Sidang

  
**Firman Noor Hasan, S.Kom., M.TI.**



**PERBANDINGAN TINGKAT AKURASI ALGORITMA NAIVE  
BAYES, *DECISION TREE* DAN *RANDOM FOREST* DALAM  
ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT  
DIGITAL NASIONAL**

**SKRIPSI**



Oleh:

**Hilmi Ammar**

**2003015162**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH PROF. DR. HAMKA  
JAKARTA  
2026**

**PERBANDINGAN TINGKAT AKURASI ALGORITMA NAIVE  
BAYES, *DECISION TREE* DAN *RANDOM FOREST* DALAM  
ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT  
DIGITAL NASIONAL**

**SKRIPSI**

Dibuat untuk Memenuhi Persyaratan Kelulusan Sarjana Teknik Informatika



Oleh:

**Hilmi Ammar**

**2003015162**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH PROF. DR. HAMKA  
JAKARTA  
2026**

## **HALAMAN PERSETUJUAN**

### **HALAMAN PERSETUJUAN**

#### **PERBANDINGAN TINGKAT AKURASI ALGORITMA NAIVE BAYES, DECISION TREE DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL**

### **SKRIPSI**

Dibuat untuk Memenuhi Persyaratan Kelulusan Sarjana Teknik

Oleh:  
Hilmi Ammar  
2003015162

Telah diperiksa dan disetujui untuk diajukan ke Sidang Ujian Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Industri dan  
Informatika UHAMKA  
Tanggal, 5 Januari 2025

Pembimbing-1



Ir. Ade Davy Wiranata, S.Kom., M.Kom.  
NIDN. 0325119302

Mengetahui,  
Ketua Program Studi Teknik Informatika



Mia Kamayani, S.T., M.T.  
NIDN. 0312028704

## HALAMAN PENGESAHAN

### HALAMAN PENGESAHAN

#### **PERBANDINGAN TINGKAT AKURASI ALGORITMA NAIVE BAYES, DECISION TREE DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL**

#### SKRIPSI

Oleh:  
Hilmi Ammar  
2003015162

Telah diuji dan dinyatakan lulus dalam Sidang Ujian Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Industri dan Informatika  
Universitas Muhammadiyah Prof. Dr. Hamka  
Tanggal, 24 April 2026

Pembimbing-1

Ir. Ade Davy Wiranata, S.Kom, M.Kom  
NIDN. 0325119302

Penguji-1

Firman Noor Hasan, S.Kom., M.Tl.  
NIDN. 0301088305

Penguji-2

Estu Sinduningrum, S.T., M.T.  
NIDN. 0314098403

Mengesahkan,  
Dekan

Fakultas Teknologi Industri dan  
Informatika

Dr. Ir. Dan Mugisidi, S.T., M.Si., IPP.  
NIDN. 0301126901

Mengetahui,  
Ketua Program Studi  
Teknik Informatika

Mia Kamayani, S.T., M.T.  
NIDN. 0312028704

## PERNYATAAN KEASLIAN

### PERNYATAAN KEASLIAN

Saya, yang membuat pernyataan

Nama : Hilmi Ammar  
NIM : 2003015162  
Judul skripsi : Perbandingan Tingkat Akurasi Algoritma *Naive Bayes*,  
*Decision Tree* dan *Random Forest* Dalam Analisis Sentimen  
Pengguna Aplikasi Samsat Digital Nasional

Menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI) dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi mana pun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, KECUALI yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Referensi.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Jakarta, 5 Mei 2026

  
Hilmi Ammar

## KATA PENGANTAR

*Assallamu'alaikum wa rohmatullahi wa barokaatuh.*

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi berjudul **“Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Universitas Muhammadiyah Prof. Dr. HAMKA.

Dalam proses penyusunannya, penulis mendapat banyak bantuan dan dukungan. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Dr. Ir. Dan Mugisidi, S.T., M.Si.,IPP selaku dekan dan Ibu Mia Kamayani Sulaeman, S.T., M.T. selaku Ketua Program Studi Fakultas Teknologi Industri dan Informatika Universitas Muhammadiyah Prof. Dr. Hamka.
2. Bapak Ir. Ade Davy Wiranata, S.Kom., M.Kom, selaku dosen pembimbing skripsi yang telah dengan sabar membimbing, memberi masukan, dan motivasi.
3. Bapak Faldy Irwiensyah, S.Kom., M.TI, selaku dosen pembimbing akademik, yang telah mendukung sejak awal perkuliahan hingga akhir masa studi.
4. Kedua orang tua penulis, Bapak Darsono dan Ibu Eli Sumarni yang telah memberikan dukungan berupa materi, nasehat serta doa.
5. Seluruh keluarga dan teman-teman tersayang yang senantiasa mendoakan serta memberikan semangat dalam menyelesaikan skripsi.

Penulis berharap skripsi ini bermanfaat bagi pembaca serta dapat menjadi referensi bagi penelitian selanjutnya dengan topik serupa. Penulis juga membuka diri terhadap saran dan masukan demi perbaikan penelitian ini di masa mendatang.

## PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

### PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Prof. DR. HAMKA (UHAMKA), saya yang bertandatangan di bawah ini:

Nama : Hilmi Ammar  
NIM : 2003015162  
Program Studi : Teknik Informatika

Menyetujui, memberikan Hak Bebas Royalti Noneksklusif (*non-exclusive royalty free right*) kepada Universitas Muhammadiyah Prof. DR. HAMKA (UHAMKA) atas karya ilmiah saya beserta perangkat yang ada (jika diperlukan) yang berjudul:

Perbandingan Tingkat Akurasi Algoritma *Naive Bayes*, *Decision Tree* dan *Random Forest* Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional

Hak Bebas Royalti Noneksklusif ini Universitas Muhammadiyah Prof. DR. HAMKA berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Jakarta, 5 Mei 2026



Hilmi Ammar

## ABSTRAK

### **Perbandingan Tingkat Akurasi Algoritma *Naive Bayes*, *Decision Tree* dan *Random Forest* Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional**

Hilmi Ammar

Penelitian ini bertujuan menganalisis sentimen pengguna terhadap aplikasi Signal – Samsat Digital Nasional menggunakan 2.000 ulasan, yang setelah melalui proses *filtering* dan *preprocessing* menjadi 1.743 data, kemudian dibagi menjadi 80% data *training* (1.394) dan 20% data *testing* (349). Data dianalisis menggunakan tiga algoritma klasifikasi, yaitu *Naive Bayes*, *Decision Tree*, dan *Random Forest*. Hasil menunjukkan bahwa *Naive Bayes* memperoleh akurasi 89.4% dengan kinerja sangat baik pada sentimen positif namun kurang optimal pada sentimen negatif, sedangkan *Decision Tree* mencapai akurasi 86.8% dengan hasil yang lebih seimbang. *Random Forest* menunjukkan performa terbaik dengan akurasi 92.9% serta kemampuan klasifikasi yang lebih seimbang antara sentimen positif dan negatif. Secara keseluruhan, *Random Forest* merupakan algoritma paling optimal dalam penelitian ini. Visualisasi data juga menunjukkan mayoritas ulasan berupa rating bintang 5, dengan distribusi sentimen 82,9% positif dan 17,1% negatif, sehingga dapat disimpulkan pengalaman pengguna aplikasi *Signal* adalah positif.

**Kata Kunci:** Analisis Sentimen, Samsat, *Naive Bayes*, *Decision Tree*, *Random Forest*

### ***Comparison of the Accuracy Levels of Naive Bayes, Decision Tree, and Random Forest Algorithms in Sentiment Analysis of National Digital Samsat Application Users***

Hilmi Ammar

*This research aims to analyze user sentiment towards the Signal – Samsat Digital Nasional application using 2,000 reviews, which after going through a filtering and preprocessing process became 1,743 data, then divided into 80% training data (1,394) and 20% testing data (349). The data were analyzed using three classification algorithms, namely Naive Bayes, Decision Tree, and Random Forest. The results show that Naive Bayes obtained an accuracy of 89.4% with very good performance on positive sentiment but less optimal on negative sentiment, while Decision Tree achieved an accuracy of 86.8% with more balanced results. Random Forest showed the best performance with an accuracy of 92.9% and a more balanced classification ability between positive and negative sentiment. Overall, Random Forest is the most optimal algorithm in this study. Data visualization also shows that the majority of reviews are in the form of 5-star ratings, with a sentiment distribution of 82.9% positive and 17.1% negative, so it can be concluded that the user experience of the Signal application is positive.*

**Keywords:** Sentiment Analysis, Samsat, *Naive Bayes*, *Decision Tree*, *Random Forest*

## DAFTAR ISI

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| HALAMAN PERSETUJUAN .....                                           | i         |
| HALAMAN PENGESAHAN .....                                            | ii        |
| PERNYATAAN KEASLIAN .....                                           | iii       |
| KATA PENGANTAR .....                                                | iv        |
| PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS ..... | v         |
| ABSTRAK .....                                                       | vi        |
| DAFTAR ISI .....                                                    | vii       |
| DAFTAR TABEL .....                                                  | ix        |
| DAFTAR GAMBAR .....                                                 | x         |
| DAFTAR NOTASI .....                                                 | xi        |
| DAFTAR LAMPIRAN .....                                               | xii       |
| <b>BAB.1 PENDAHULUAN .....</b>                                      | <b>1</b>  |
| 1.1 Latar Belakang .....                                            | 1         |
| 1.2 Rumusan Masalah .....                                           | 3         |
| 1.3 Batasan Masalah .....                                           | 4         |
| 1.4 Tujuan Penelitian .....                                         | 4         |
| 1.5 Manfaat Penelitian .....                                        | 5         |
| 1.6 Sistematika Penulisan .....                                     | 6         |
| <b>BAB.2 DASAR TEORI .....</b>                                      | <b>8</b>  |
| 2.1 Kerangka Teori .....                                            | 8         |
| 2.2 Penelitian Terdahulu .....                                      | 8         |
| 2.3 Jenis Referensi .....                                           | 11        |
| 2.3.1 Analisis Sentimen .....                                       | 11        |
| 2.3.2 Signal – Samsat Digital Nasional .....                        | 11        |
| 2.3.3 <i>Scraping</i> Data .....                                    | 12        |
| 2.3.4 <i>Naïve Bayes</i> .....                                      | 12        |
| 2.3.5 <i>Decision Tree</i> .....                                    | 13        |
| 2.3.6 <i>Random Forest</i> .....                                    | 13        |
| 2.3.7 Pra-Pemrosesan Teks .....                                     | 13        |
| 2.3.8 <i>TF-IDF</i> .....                                           | 14        |
| 2.3.9 <i>Confusion Matrix</i> .....                                 | 15        |
| <b>BAB.3 METODOLOGI PENELITIAN .....</b>                            | <b>16</b> |
| 3.1 Alur Penelitian .....                                           | 16        |
| 3.1.1 Pengumpulan Data .....                                        | 16        |
| 3.1.2 <i>Filter</i> Data .....                                      | 17        |
| 3.1.3 <i>Text Processing</i> .....                                  | 17        |
| 3.1.4 Pelabelan Data .....                                          | 20        |

|                                                |           |
|------------------------------------------------|-----------|
| 3.1.5 <i>Split</i> Data.....                   | 21        |
| 3.1.6 Pembobotan <i>TF-IDF</i> .....           | 21        |
| 3.1.7 Klasifikasi Model.....                   | 22        |
| 3.1.8 Evaluasi Algoritma.....                  | 22        |
| 3.1.9 Visualisasi Data.....                    | 23        |
| 3.2 Identifikasi Kebutuhan .....               | 23        |
| 3.3 Jadwal Penelitian.....                     | 24        |
| <b>BAB.4 HASIL DAN PEMBAHASAN .....</b>        | <b>25</b> |
| 4.1 Pengumpulan Data .....                     | 25        |
| 4.2 Filter Data.....                           | 26        |
| 4.3 <i>Text Processing</i> .....               | 28        |
| 4.4 Pelabelan Data.....                        | 29        |
| 4.5 <i>Split</i> Data.....                     | 31        |
| 4.6 Pembobotan <i>TF-IDF</i> .....             | 32        |
| 4.7 Klasifikasi Model.....                     | 34        |
| 4.8 Evaluasi Algoritma.....                    | 35        |
| 4.9 Hasil Perbandingan Akurasi Algoritma ..... | 40        |
| 4.10 Visualisasi Data.....                     | 41        |
| <b>BAB.5 KESIMPULAN DAN SARAN .....</b>        | <b>44</b> |
| 5.1 Kesimpulan .....                           | 44        |
| 5.2 Saran.....                                 | 45        |
| <b>DATAR PUSTAKA .....</b>                     | <b>46</b> |
| <b>LAMPIRAN.....</b>                           | <b>52</b> |

## DAFTAR TABEL

|                                                         |    |
|---------------------------------------------------------|----|
| Tabel 2. 1. Penelitian Terdahulu .....                  | 8  |
| Tabel 2.2. Nilai Aktual <i>Confusion Matrix</i> .....   | 15 |
| Tabel 3.1. Spesifikasi <i>Hardware</i> .....            | 23 |
| Tabel 3.2. Spesifikasi <i>Software</i> .....            | 24 |
| Tabel 3.3. Jadwal Rencana Kegiatan Penelitian .....     | 24 |
| Tabel 4.1. Ilustrasi Hasil <i>Web Scraping</i> .....    | 26 |
| Tabel 4.2. Ilustrasi Hasil <i>Text Processing</i> ..... | 29 |
| Tabel 4.3. Ilustrasi Hasil Pelabelan Sentimen.....      | 30 |
| Tabel 4.4. Ilustrasi Hasil Bobot Kata TF-IDF.....       | 33 |
| Tabel 4.5. Hasil Perbandingan Evaluasi Algoritma .....  | 40 |

## DAFTAR GAMBAR

|                                                                        |    |
|------------------------------------------------------------------------|----|
| Gambar 2.1. Kerangka Teori.....                                        | 8  |
| Gambar 3.1. Alur Penelitian.....                                       | 16 |
| Gambar 4.1. Alur Pengumpulan Data .....                                | 25 |
| Gambar 4.2. Hasil Pengumpulan Data Melalui <i>Web Scraping</i> .....   | 25 |
| Gambar 4.3. <i>Filtering</i> Kolom Utama Penelitian.....               | 27 |
| Gambar 4.4. Hasil Proses <i>Filtering</i> Data .....                   | 28 |
| Gambar 4.5. Hasil Proses <i>Text Processing</i> .....                  | 28 |
| Gambar 4.6. Hasil Pelabelan Data Sentimen .....                        | 30 |
| Gambar 4.7. Hasil <i>Split</i> Data .....                              | 32 |
| Gambar 4.8. Hasil Pembobotan <i>TF-IDF</i> .....                       | 34 |
| Gambar 4.9. Hasil Klasifikasi Model.....                               | 35 |
| Gambar 4.10. Hasil <i>Confusion Matrix Heatmap Naïve Bayes</i> .....   | 36 |
| Gambar 4.11. Hasil <i>Confusion Matrix Heatmap Decision Tree</i> ..... | 37 |
| Gambar 4.12. Hasil <i>Confusion Matrix Heatmap Random Forest</i> ..... | 39 |
| Gambar 4.13. Hasil Distribusi Sentimen Dengan Kamus Lexicon.....       | 41 |
| Gambar 4.14. Hasil Jumlah Pengguna Memberikan Rating.....              | 42 |
| Gambar 4.15. Hasil <i>Word Cloud</i> Kata .....                        | 43 |

## DAFTAR NOTASI

| No. | Notasi                          | Keterangan                                                                                 |
|-----|---------------------------------|--------------------------------------------------------------------------------------------|
| 1   | TP<br>( <i>True Positive</i> )  | Jumlah data dengan nilai aktual kelas positif yang juga diprediksi sebagai kelas positif.  |
| 2   | TN<br>( <i>True Negative</i> )  | Jumlah data dengan nilai aktual kelas negatif yang juga diprediksi sebagai kelas negatif.  |
| 3   | FP<br>( <i>False Positive</i> ) | Jumlah data dengan nilai aktual kelas positif yang salah diprediksi sebagai kelas negatif. |
| 4   | FN<br>( <i>False Negative</i> ) | Jumlah data dengan nilai aktual kelas positif yang salah diprediksi sebagai kelas negatif. |

## DAFTAR LAMPIRAN

|                                           |    |
|-------------------------------------------|----|
| Lampiran 1. LoA.....                      | 52 |
| Lampiran 2. Editor Decision .....         | 53 |
| Lampiran 3. Sertifikat Publisher .....    | 54 |
| Lampiran 4. Lembar Bimbingan Skripsi..... | 55 |
| Lampiran 5. Turnitin .....                | 56 |
| Lampiran 6. Lembar Revisi Penguji 1 ..... | 57 |
| Lampiran 7. Lembar Revisi Penguji 2 ..... | 58 |
| Lampiran 8. Jurnal .....                  | 74 |

## DATAR PUSTAKA

- Afandi Afandi, Zainal fatah, Sri Roekminiati, & Ika Devy Pramudiana. (2025). Inovasi Layanan Publik dalam Era Digital: Tinjauan Persepsi Generasi Muda pada Samsat Digital di Lumajang. *SOSIAL : Jurnal Ilmiah Pendidikan IPS*, 3(3), 265–289. <https://doi.org/10.62383/sosial.v3i3.1064>
- Afifudin, A., Kadir, A., & Isnaini, I. (2020). Implementasi Kebijakan Sistem Administrasi Manunggal Satu Atap di Kabupaten Mandailing Natal. *Strukturasi: Jurnal Ilmiah Magister Administrasi Publik*, 2(1), 59–69. <https://doi.org/10.31289/strukturasi.v2i1.43>
- Alun Sujjadaa, Somantri, Juwita Nurfazri Novianti, & Indra Griha Tofik Isa. (2023). ANALISIS SENTIMEN TERHADAP REVIEW BANK DIGITAL PADA GOOGLE PLAY STORE MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM). *Jurnal Rekayasa Teknologi Nusa Putra*, 9(2), 122–135. <https://doi.org/10.52005/rekayasa.v9i2.345>
- Amal, M. I., Rahmasita, E. S., Suryaputra, E., & Rakhmawati, N. A. (2022). Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(3). <https://doi.org/10.28932/jutisi.v8i3.5483>
- Aryandra, C. A., Sanggrahita, D., Kusumaningrum, S. D., Ahyubi, I. Al, & Iksan, N. (2025). PENDEKATAN BERBASIS DECISION TREE DENGAN POLA WILDCARD UNTUK MENGIDENTIFIKASI PLAT NOMOR KENDARAAN PULAU JAWA. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3). <https://doi.org/10.23960/jitet.v13i3.6728>
- Br Sinulingga, J. E., & Sitorus, H. C. K. (2024). Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, 14(1), 42–53. <https://doi.org/10.34010/jamika.v14i1.11946>

- Bustami, B. (2013). PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASI DATA NASABAH ASURANSI. *TECHSI - Jurnal Teknik Informatika*, 5(2). <https://doi.org/10.29103/techsi.v5i2.154>
- Chandra, R., & Sipayung, E. M. (2025). Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Menggunakan Algoritma Naive Bayes Classifier. *Jurnal Nasional Teknologi dan Sistem Informasi*, 10(3), 156–164. <https://doi.org/10.25077/TEKNOSI.v10i3.2024.156-164>
- Fajri. (t.t.). *Lexicon InSet*. Diambil 5 Maret 2025, dari <https://github.com/fajri91/InSet>
- Gifari, O. I., Adha, Muh., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40. <https://doi.org/10.46229/jifotech.v2i1.330>
- Hananto, A. D., Erfiana, A. M., Putri, B. L. P., Putri, P. D., & Kurniawan, F. (2023). Algoritma Machine Learning Naïve Bayes pada Analisis Sentimen Kesepakatan Polri dan GNPF-MUI pada Aksi Bela Islam III “212.” *SINTA Journal (Science, Technology, and Agricultural)*, 4(2), 151–160. <https://doi.org/10.37638/sinta.4.2.151-160>
- Hasan, I. K., Resmawan, R., & Ibrahim, J. (2022). Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa. *Indonesian Journal of Applied Statistics*, 5(1), 58. <https://doi.org/10.13057/ijas.v5i1.58056>
- Hidayah, L., & Rosadi, M. I. (2024). PENERAPAN ALGORITMA RANDOM FOREST UNTUK MEMREDIKSI JUMLAH SANTRI BARU. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3S1). <https://doi.org/10.23960/jitet.v12i3S1.5237>
- Huda, D. N. I., Prianto, C., & Awangga, R. M. (2023). ANALISIS SENTIMEN PERBANDINGAN LAYANAN JASA PENGIRIMAN KURIR PADA ULASAN PLAY STORE MENGGUNAKAN METODE DECISION TREE

- DAN RANDOM FOREST. *JURNAL ILMIAH INFORMATIKA*, 11(02), 150–158. <https://doi.org/10.33884/jif.v11i02.7952>
- Irfani, F. F. (2020). ANALISIS SENTIMEN REVIEW APLIKASI RUANGGURU MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *JBMI (Jurnal Bisnis, Manajemen, dan Informatika)*, 16(3), 258–266. <https://doi.org/10.26487/jbmi.v16i3.8607>
- Julianti, O. N., Suarna, N., & Prihartono, W. (2024). PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANALISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2936–2941. <https://doi.org/10.36040/jati.v8i3.9613>
- Kacung, S., Pamungkas Putra Bagyana, C., & Cahyono, D. (2024). ANALISIS SENTIMEN TERHADAP LAYANAN SAMSAT DIGITAL NASIONAL (SIGNAL) MENGGUNAKAN METODE SVM. *Jurnal Mnemonic*, 7(1), 118–122. <https://doi.org/10.36040/mnemonic.v7i1.9557>
- Kharisma, I. L., Septiani, D. A., Fergina, A., & Kamdan, K. (2023). Penerapan Algoritma Decision Tree untuk Ulasan Aplikasi Vidio di Google Play. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(2), 218–226. <https://doi.org/10.25077/TEKNOSI.v9i2.2023.218-226>
- Koto, F., & Rahmaningtyas, G. Y. (2017). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *2017 International Conference on Asian Language Processing (IALP)*, 391–394. <https://doi.org/10.1109/IALP.2017.8300625>
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(9), 4305–4313. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
- Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. *Journal of Information Systems for Public Health*, 6(1), 44. <https://doi.org/10.22146/jisph.44455>

- Mardianto, R., Stefanie Quinevera, & Rochimah, S. (2024a). Perbandingan Metode Random Forest, Convolutional Neural Network, dan Support Vector Machine Untuk Klasifikasi Jenis Mangga. *Journal of Applied Computer Science and Technology*, 5(1), 63–71. <https://doi.org/10.52158/jacost.v5i1.742>
- Mardianto, R., Stefanie Quinevera, & Rochimah, S. (2024b). Perbandingan Metode Random Forest, Convolutional Neural Network, dan Support Vector Machine Untuk Klasifikasi Jenis Mangga. *Journal of Applied Computer Science and Technology*, 5(1), 63–71. <https://doi.org/10.52158/jacost.v5i1.742>
- Merinda Lestandy, Abdurrahim Abdurrahim, & Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(4), 802–808. <https://doi.org/10.29207/resti.v5i4.3308>
- Musthafa, A., Harmini, T., Rafiq, A., & Marantika, N. (2025). Pemanfaatan Machine Learning dalam Menganalisis Sentimen Terhadap Program TAPERA di Platform Digital X. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 587–597. <https://doi.org/10.57152/malcom.v5i2.1801>
- Nanda, S., Muallfah, D., & Fitri, D. A. (2022). Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest. *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, 3(2), 210–219. <https://doi.org/10.31102/jatim.v3i2.1592>
- Premana Putra, I. G. B., Sudarma, M., & Manuaba, I. B. G. (2023). Penerapan Metode Extreme Programming pada Rancang Bangun Sistem Analisis Sentimen Portal Berita. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(6), 1369–1378. <https://doi.org/10.25126/jtiik.2023106904>
- Rahman, I. F., Hasanah, A. N., & Heryana, N. (2024). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SAMSAT DIGITAL NASIONAL (SIGNAL) DENGAN MENGGUNAKAN METODE NAÏVE BAYES

- CLASSIFIER. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2).  
<https://doi.org/10.23960/jitet.v12i2.4073>
- Rania, N. Z., & Syah, R. D. (2024). ANALISIS SENTIMEN TERHADAP APLIKASI GOJEK PADA PLAY STORE MENGGUNAKAN METODE RANDOM FOREST CLASSIFIER. *Jurnal Ilmiah Informatika Komputer*, 29(2), 144–153. <https://doi.org/10.35760/ik.2024.v29i2.11877>
- Rizaldi, S. A. R., Alam, S., & Kurniawan, I. (2023). Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, 2(3), 109–117. <https://doi.org/10.55123/storage.v2i3.2334>
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(4), 377–383. <https://doi.org/10.47233/jteksis.v5i4.913>
- Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmadden, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19. <https://doi.org/10.57152/malcom.v3i1.591>
- Siki, Y. C. H., Tokan, T. B. A., Manehat, D. J., & Mau, S. D. B. (2025). Perbandingan Metode Naïve Bayes dan K-Nearest Neighbor Terhadap Sentimen Analisis Pinjaman Online. *Jurnal Media ...*, 6, 1614–1622. <https://doi.org/https://doi.org/10.55338/jumin.v6i3.5687>
- Suriyani BB, La Ode Agus Said, & Astriana Syaqlra. (2024). DIGITALISASI LAYANAN SAMSAT: Peningkatan Kepatuhan Pajak Kendaraan Bermotor Melalui e-Samsat di Kabupaten Muna. *Journal Publicuho*, 7(4), 2111–2126. <https://doi.org/10.35817/publicuho.v7i4.573>
- Syah, H., & Witanti, A. (2022). ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

- (SVM). *Jurnal Sistem Informasi dan Informatika (Simika)*, 5(1), 59–67.  
<https://doi.org/10.47080/simika.v5i1.1411>
- Triyosanti, C., Mufarokhah, D. F., Angelina, S. N., Cahyono, A. S., & Margayaningsih, D. I. (2024). DAMPAK PENERAPAN APLIKASI LAYANAN SIGNAL (SAMSAT DIGITAL NASIONAL) TERHADAP PELAYANAN PAJAK KENDARAAN BERMOTOR DI TULUNGAGUNG. *JSPH: Jurnal Sosial Politik Humaniora*, 1(3), 42–49.  
<https://doi.org/10.59966/jsph.v1i3.1426>
- Ulfa, S. N., & Frinaldi, A. (2025). INOVASI PEMERINTAH DAERAH DALAM PENERAPAN KEBIJAKAN PUBLIK DAN PELAYANAN PUBLIK. *SOCIAL: Jurnal Inovasi Pendidikan IPS*, 5(2), 610–621.  
<https://doi.org/10.51878/social.v5i2.5965>
- Ulfah, A., & Najiah, I. (2023). IMPLEMENTASI WEB SCRAPING PADA SITUS JURNAL SINTA MENGGUNAKAN FRAMEWORK SELENIUM WEBDRIVER PYTHON. *JIKA (Jurnal Informatika)*, 7(1), 29.  
<https://doi.org/10.31000/jika.v7i1.7037>
- Wijaya, D., Saputra, R. A., & Irwiensyah, F. (2024). Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(4), 2369–2380. <https://doi.org/10.30865/klik.v4i4.1738>

## LAMPIRAN

### 1. LoA (Lecture of Acceptance) Jurnal

eISSN: 2598-5981, pISSN: 2548-3846

**INDONESIAN JOURNAL OF APPLIED INFORMATICS IJAI**

No : 0057/IJAI-Acept/VI/2025  
Hal : *Accepted Artikel*

Kepada  
Hilmi Ammar  
Author IJAI

Dengan hormat,

Bersama surat ini, kami Editor Indonesian Journal of Applied Informatics menginformasikan kepada Bapak/Ibu Author bahwa naskah:

**ID Artikel : 111010**

**Judul :**

“Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional”

**Author :**

Hilmi Ammar, Ade Davy Wiranata  
Email koresponding : hilmixammar@gmail.com

Telah melalui tahap review sehingga dinyatakan **DITERIMA**. Hasil revisi artikel yang siap terbit akan kami terbitkan pada IJAI (*Indonesian Journal of Applied Informatics*) Terbitan Vol. 10 No. 1.

Demikian atas perhatiannya kami sampaikan terima kasih.

Surakarta, 31 Desember 2025

Editor in Chief IJAI



Fendi Aji Purnomo, S.Si., M.Eng.

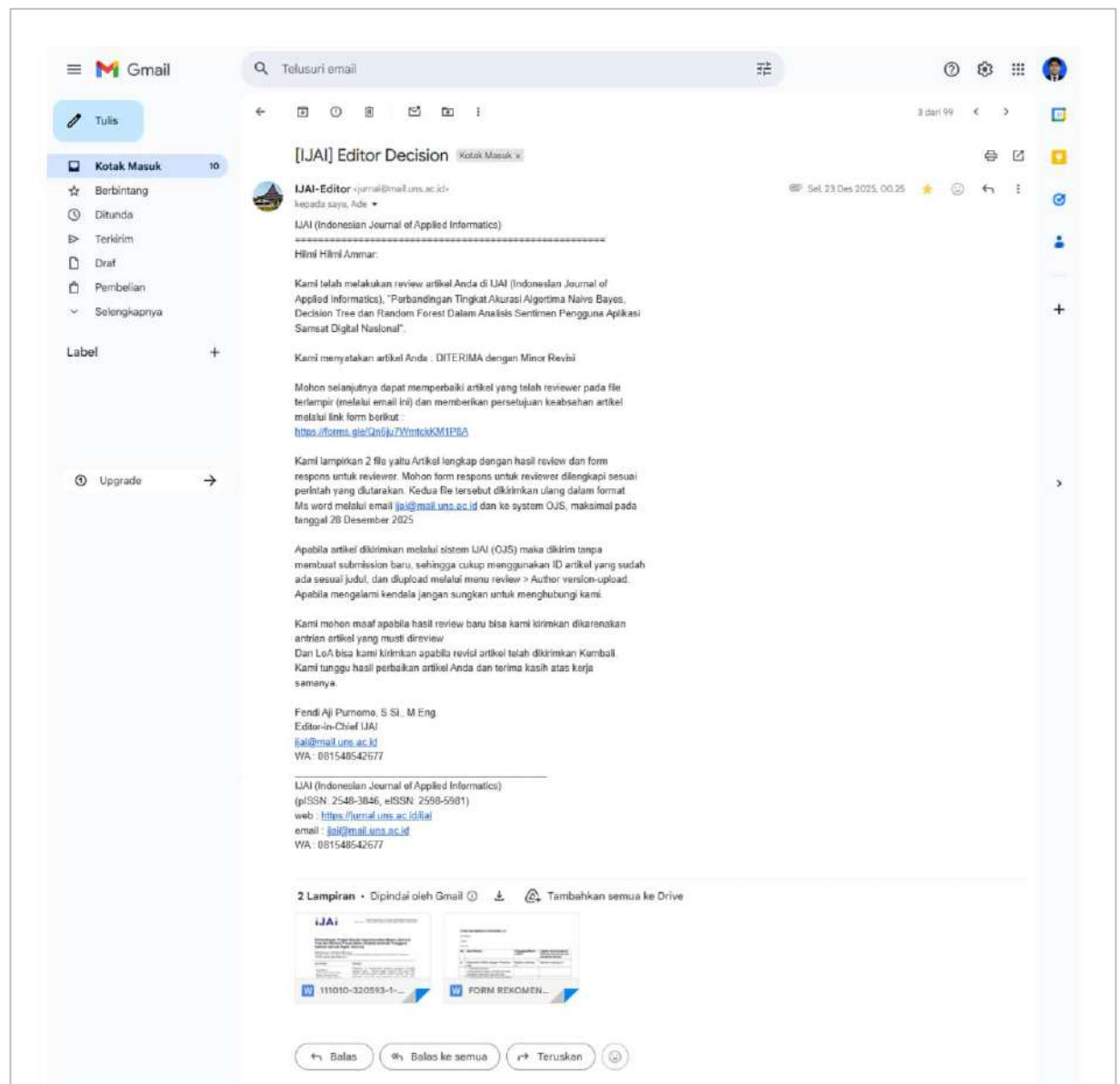
---

**IJAI (Indonesian Journal of Applied Informatics)**

Email : [ijai@mipa.uns.ac.id](mailto:ijai@mipa.uns.ac.id)

Website : [journal.uns.ac.id/ijai](http://journal.uns.ac.id/ijai)

## 2. Editor Decision



Lampiran 2. Editor Decision

### 3. Sertifikat Publisher



Lampiran 3. Sertifikat Publisher

#### 4. Lembar Bimbingan Skripsi



UNIVERSITAS MUHAMMADIYAH PROF. DR. HAMKA  
**FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA**  
 Jl. Tanah Merdeka No. 6 Kp. Rambutan, Ciracas, Jakarta Timur. Telp. (021) 8400941: Fax. (021) 87782739  
 Website : ft.uhamka.ac.id: E-mail : ftl@uhamka.ac.id

### LEMBAR BIMBINGAN SKRIPSI

Nama : Hilmi Ammar  
 NIM : 2003015162  
 Program Studi : Teknik Informatika  
 Judul : Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree, Random Forest Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional

Dosen Pembimbing : Ir. Ade Davy Wiranata, S.Kom., M.Kom

| No. | Tanggal    | Keterangan                                                            | Paraf |
|-----|------------|-----------------------------------------------------------------------|-------|
| 1   | 18/10/2023 | Pengajuan Judul                                                       |       |
| 2   | 17/11/2023 | Bimbingan Bab 1 dan Pembahasan Turnitin                               |       |
| 3   | 28/12/2023 | Bimbingan Bab 2 dan Pembahasan Jurnal                                 |       |
| 4   | 11/01/2024 | Bimbingan Bab 1-3 dan Pengumpulan Untuk Seminar Proposal              |       |
| 5   | 17/01/2024 | Acc Persetujuan Seminar Proposal                                      |       |
| 6   | 09/10/2025 | Pergantian Judul dan Pembahasan Jurnal                                |       |
| 7   | 27/10/2025 | Bimbingan Terkait Metadata Publikasi Jurnal                           |       |
| 8   | 02/01/2026 | Bimbingan Bab 4-5 dan Pembahasan Revisi Serta Penerimaan Jurnal & LoA |       |
| 9   | 05/01/2026 | Acc Laporan Skripsi                                                   |       |

Mengetahui,  
 Dosen Pembimbing

Ir. Ade Davy Wiranata, S.Kom., M.Kom  
 NIDN. 0325119302

Mahasiswa

Hilmi Ammar  
 NIM. 2003015162

## 5. Turnitin


Page 1 of 21 - Cover Page
Submission ID: trn:oid::1:3381922320

# 52 52

## HILMI AMMAR - Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest Dalam Analisis...

29/9/2025  
FTII 2  
Universitas Muhammadiyah Prof. Dr. Hamka

### Document Details

|                                                              |                                              |
|--------------------------------------------------------------|----------------------------------------------|
| Submission ID<br>trn:oid::1:3381922320                       | 15 Pages<br>4,514 Words<br>28,899 Characters |
| Submission Date<br>Oct 22, 2025, 10:46 AM GMT+7              |                                              |
| Download Date<br>Oct 22, 2025, 1:01 PM GMT+7                 |                                              |
| File Name<br>HILMI_AMMAR_TI_TURNITIN_KE-3_-_HILMI_AMMAR.docx |                                              |
| File Size<br>5.6 MB                                          |                                              |

### 16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

#### Filtered from the Report

- Bibliography

#### Exclusions

- 2 Excluded Matches

#### Top Sources

|     |                                  |
|-----|----------------------------------|
| 9%  | Internet sources                 |
| 12% | Publications                     |
| 4%  | Submitted works (Student Papers) |

#### Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.


Page 2 of 21 - Integrity Overview
Submission ID: trn:oid::1:3381922320

Lampiran 5. Turnitin

## 6. Lembar Revisi – Penguji 1

**LEMBAR REVISI SIDANG SKRIPSI TEKNIK INFORMATIKA (PENGUJI-1)**  
**FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA UHAMKA**

|                         |                                                                                                                                                         |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Nama Mahasiswa          | : Hilmi Ammar                                                                                                                                           |
| NIM                     | : 2003015162                                                                                                                                            |
| Hari Tanggal Sidang     | : Sabtu, 25 April 2026                                                                                                                                  |
| Nama Pembimbing Skripsi | : Ir. Ade Davy Wiranata, S.Kom., M.Kom.                                                                                                                 |
| Judul Skripsi           | : Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional |

**Catatan**

|   |                                                                                   |
|---|-----------------------------------------------------------------------------------|
| 1 | Buat versi skripsi dengan penulisan yang sesuai format panduan penulisan skripsi. |
| 2 | Harus didalami kembali terkait topik penelitian Anda.                             |
| 3 |                                                                                   |
| 4 |                                                                                   |
| 5 |                                                                                   |
| 6 |                                                                                   |

| VALIDASI REVISI             | NAMA DOSEN                            | TANGGAL REVISI | PARAF                                                                                 |
|-----------------------------|---------------------------------------|----------------|---------------------------------------------------------------------------------------|
| (Ketua Sidang)<br>Penguji-1 | Firman Noor Hasan, S.Kom., M.Tl.      | 02/05/2026     |  |
| Penguji-2                   | Estu Sinduningrum, S.T., M.T.         | 4/5/2026       |  |
| Pembimbing-1                | Ir. Ade Davy Wiranata, S.Kom., M.Kom. | 30/04/26       |  |
| Pembimbing-2                |                                       |                |                                                                                       |

Selanjutnya, yang bersangkutan harus segera menyelesaikan permasalahan sehubungan dengan skripsi ini, selambat-lambatnya 10 (Sepuluh) hari setelah tanggal pelaksanaan sidang.

- 0 Apabila revisi telah selesai dan mendapatkan approval (penguji, pembimbing, Kaprodi dan Dekan), maka tulisan (Skripsi, Jurnal) dan Program dikumpulkan di spss.uhamka.ac.id atau Google Drive (Fakultas/Perpustakaan, Pembimbing dan Program Studi)
- 0 Berkas disusun sesuai petunjuk dan tanda tangan setiap berkas Asli
- 0 Batas Akhir Revisi 4 Mei 2026 (hh/bb/ttt)
- 0 Batas Akhir Pengumpulan Berkas Skripsi dan Jurnal/Resume 5 Mei 2026 (hh/bb/ttt)

*Wassalamu'alaikum wa Rahmatullahi wa Barokaatuh,*

**Mia Kamayani, Sulaeman, S.T., M.T.**  
Ketua Program Studi Teknik Informatika

Catatan: Daftar revisi ini diserahkan kepada mahasiswa untuk acuan revisi bagi Dosen Pembimbing

Lampiran 6. Lembar Revisi Penguji 1

## 7. Lembar Revisi – Penguji 2

**LEMBAR REVISI SIDANG SKRIPSI TEKNIK INFORMATIKA (PENGUJI-2)**  
**FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA UHAMKA**

|                         |                                                                                                                                                         |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Nama Mahasiswa          | : Hilmi Ammar                                                                                                                                           |
| NIM                     | : 2003015162                                                                                                                                            |
| Hari Tanggal Sidang     | : Sabtu, 25 April 2026                                                                                                                                  |
| Nama Pembimbing Skripsi | : Ir. Ade Davy Wiranata, S.Kom., M.Kom.                                                                                                                 |
| Judul Skripsi           | : Perbandingan Tingkat Akurasi Algoritma Naive Bayes, Decision Tree dan Random Forest Dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional |

**Catatan**

|   |                                                                                        |
|---|----------------------------------------------------------------------------------------|
| 1 | Perbaiki bab 1 batasan masalah raw data di ambil dari mana, jumlah & mengapa 80% & 20% |
| 2 | bab 2 untuk modifikasi atau ketebaharuan penelitian harus di jelaskan                  |
| 3 |                                                                                        |
| 4 |                                                                                        |
| 5 |                                                                                        |
| 6 |                                                                                        |

| VALIDASI REVISI             | NAMA DOSEN                            | TANGGAL REVISI | PARAF                                                                                 |
|-----------------------------|---------------------------------------|----------------|---------------------------------------------------------------------------------------|
| (Ketua Sidang)<br>Penguji-1 | Firman Noor Hasan, S.Kom., M.TI.      | 02/05/2026     |  |
| Penguji-2                   | Estu Sinduningrum, S.T., M.T.         | 4/5/2026       |  |
| Pembimbing-1                | Ir. Ade Davy Wiranata, S.Kom., M.Kom. | 30/04/26       |  |
| Pembimbing-2                |                                       |                |                                                                                       |

Selanjutnya, yang bersangkutan harus segera menyelesaikan permasalahan sehubungan dengan skripsi ini, selambat-lambatnya 10 (Sepuluh) hari setelah tanggal pelaksanaan sidang.

- 0 Apabila revisi telah selesai dan mendapatkan approval (penguji, pembimbing, Kaprodi dan Dekan), maka tulisan (Skripsi, Jurnal) dan Program dikumpulkan di spss.uhamka.ac.id atau Google Drive (Fakultas/Perpustakaan, Pembimbing dan Program Studi)
- 0 Berkas disusun sesuai petunjuk dan tanda tangan setiap berkas Asli
- 0 Batas Akhir Revisi 4 Mei 2026 (hh/bb/tttt)
- 0 Batas Akhir Pengumpulan Berkas Skripsi dan Jurnal/Resume 5 Mei 2026 (hh/bb/tttt)

*Wassalamu'alaikum wa Rahmatullahi wa Barakaatuh,*

**Mia Kamayani, Sulaeman, S.T., M.T.**  
Ketua Program Studi Teknik Informatika

Catatan: Daftar revisi ini diserahkan kepada mahasiswa untuk acuan revisi bagi Dosen Pembimbing

Lampiran 7. Lembar Revisi Penguji 2

## 8. Jurnal



IJAI (Indonesian Journal of Applied Informatics)  
Vol. 10 No. 1 Tahun 2025 pISSN: 2548-3846, eISSN: 2598-5981

### Perbandingan Tingkat Akurasi Algoritma *Naïve Bayes*, *Decision Tree* dan *Random Forest* dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional

Hilmi Ammar<sup>1</sup>, Ade Davy Wiranata<sup>1\*</sup>

<sup>1</sup>Program Studi S1 Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia

\*Email: adedavy@uhamka.ac.id

#### Info Artikel

##### Kata Kunci :

Analisis Sentimen, *Naïve Bayes*,  
*Decision Tree*, *Random Forest*,  
Samsat Digital Nasional

##### Keywords :

*Sentiment Analysis*, *Naïve Bayes*,  
*Decision Tree*, *Random Forest*,  
*National Digital Samsat*

##### Tanggal Artikel

Dikirim : 13 November 2025

Direvisi : 25 Desember 2025

Diterima : 30 Desember 2025

#### Abstrak

Penelitian ini menganalisis sentimen pengguna terhadap aplikasi Signal – Samsat Digital Nasional. Dari total 2.000 ulasan yang terkumpul, setelah melalui tahapan *filter* data diperoleh 1.743 data yang digunakan untuk analisis. Data tersebut kemudian dipecah menjadi 1.394 data pelatihan (alokasi 80%) dan 349 data pengujian (alokasi 20%). Tiga model klasifikasi yang digunakan dalam penelitian ini meliputi *Naïve Bayes*, *Decision Tree*, dan *Random Forest*. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki performa terbaik dengan akurasi 92,9% serta keseimbangan tinggi dalam mengenali sentimen positif (*f1-score* 95,9%) dan negatif (*f1-score* 73,1%). *Naïve Bayes* mencapai akurasi 89,4% namun kurang seimbang, sedangkan *Decision Tree* memperoleh akurasi 86,8% dengan hasil yang lebih stabil dibandingkan *Naïve Bayes*. Secara keseluruhan, visualisasi data berdasarkan analisis menggunakan kamus *lexicon* menunjukkan bahwa 82,9% ulasan bersentimen positif dan didominasi oleh rating bintang lima, yang mengindikasikan bahwa pengalaman pengguna terhadap aplikasi Signal – Samsat Digital Nasional tergolong sangat baik.

#### Abstract

This research analyzes user sentiment towards the Signal – National Digital Samsat application. From a total of 2,000 collected reviews, after going through the data filtering stage, 1,743 data were obtained for analysis. The data was then split into 1,394 training data (80% allocation) and 349 testing data (20% allocation). Three classification models used in this study were *Naïve Bayes*, *Decision Tree*, and *Random Forest*. The results showed that *Random Forest* had the best performance with 92.9% accuracy and high balance in recognizing positive (*f1-score* 95.9%) and negative (*f1-score* 73.1%) sentiment. *Naïve Bayes* achieved 89.4% accuracy but was less balanced, while *Decision Tree* achieved 86.8% accuracy with more stable results than *Naïve Bayes*. Overall, data visualization based on analysis using the *lexicon* dictionary shows that 82.9% of reviews are positive and dominated by five-star ratings, which indicates that the user experience of the Signal – Samsat Digital Nasional application is classified as very good.

## 1. PENDAHULUAN

Era digitalisasi telah mengubah paradigma layanan publik di Indonesia, salah satunya adalah adopsi teknologi digital dalam berbagai sektor pemerintahan [1]. Pada sektor pemerintahan, transformasi digital berperan penting dalam menciptakan sistem pelayanan publik yang lebih efisien dan transparan serta membuka ruang partisipasi masyarakat dalam proses pengambilan keputusan. Sejalan dengan perkembangan teknologi yang semakin pesat, pemerintah terus berupaya meningkatkan kualitas serta efektivitas layanan publik melalui penerapan berbagai inovasi digital dalam tata kelola pemerintahan [2].

Merujuk pada Undang-Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, Sistem Administrasi Manunggal Satu Atap (SAMSAT) merupakan mekanisme administrasi terpadu yang diterapkan di Indonesia untuk mengelola administrasi perpajakan dan penerimaan negara terkait kendaraan bermotor. Tujuan dari penerapan SAMSAT adalah meningkatkan kepatuhan wajib pajak kendaraan, sekaligus menyediakan layanan administrasi yang efektif dan terintegrasi [3]. Melalui sistem ini, pemerintah dapat memantau penerimaan pajak kendaraan bermotor guna mendukung pembiayaan pembangunan infrastruktur dan penyediaan layanan publik lainnya [4].

Dalam upaya meningkatkan kualitas layanan administratif, SAMSAT mengembangkan inovasi melalui peluncuran aplikasi Samsat Digital Nasional (SIGNAL), yang dapat diunduh secara resmi melalui platform *Google Play Store*. Aplikasi ini dikembangkan melalui kerja sama antara Korlantas Polri, Kementerian Dalam Negeri dan Jasa Raharja dengan tujuan menyediakan layanan pembayaran Pajak Kendaraan Bermotor (PKB) serta pengesahan tahunan Surat Tanda Nomor Kendaraan (STNK) secara *online* [5]. Tujuan pengembangan aplikasi SIGNAL adalah untuk menyederhanakan proses birokrasi, mengurangi antrean di kantor SAMSAT, meningkatkan kepatuhan wajib pajak, serta meminimalisir potensi praktik pungutan liar melalui sistem yang lebih transparan [6].

Berdasarkan ulasan pengguna di *Google Play Store*, aplikasi Samsat Digital Nasional (SIGNAL) diapresiasi karena kemudahan pembayaran pajak kendaraan, namun masih dinilai kurang stabil. Beberapa pengguna melaporkan kendala seperti proses verifikasi yang sulit, kesalahan transaksi, dan waktu pemrosesan yang lambat. Hal ini menunjukkan bahwa keberhasilan aplikasi tidak hanya ditentukan oleh aspek teknis, tetapi juga oleh kemampuannya dalam memenuhi kebutuhan dan harapan pengguna. Oleh karena itu, evaluasi pengalaman pengguna melalui analisis sentimen ulasan diperlukan untuk mengidentifikasi faktor kepuasan dan ketidakpuasan sebagai dasar perbaikan dan pengembangan aplikasi di masa mendatang [7].

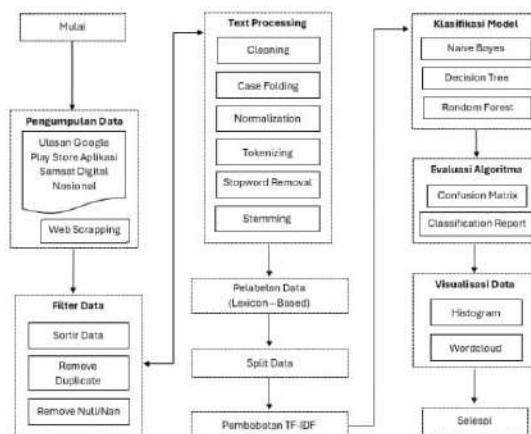
Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi opini, emosi, dan tanggapan pengguna dari data tekstual [8]. Analisis ini memanfaatkan teknik pengolahan bahasa dan algoritma *machine learning* untuk mengklasifikasikan ulasan ke dalam tiga kategori, yaitu sentimen positif, negatif, dan netral [9]. Beberapa algoritma yang umum digunakan antara lain *Naive Bayes*, *Decision Tree*, dan *Random Forest*, di mana masing-masing memiliki keunggulan tersendiri. Namun, berbagai penelitian menunjukkan bahwa tidak ada algoritma yang selalu unggul pada semua kondisi, karena kinerjanya sangat dipengaruhi oleh karakteristik dataset seperti jumlah data, fitur dan penggunaan bahasa [10].

Penelitian terkait analisis sentimen pada aplikasi Samsat Digital Nasional (SIGNAL) telah banyak dilakukan sebelumnya. Imam Fathur Rahman menggunakan algoritma *Naive Bayes* dan memperoleh akurasi 89% dengan nilai *recall* 85% dan *f1-score* 87% dari 1.000 data ulasan [11]. Penelitian lain oleh Slamet Kacung menerapkan algoritma *Support Vector Machine* (SVM) dan menghasilkan akurasi 96,19% dengan rata-rata *K-Fold* sebesar 97,65% [12]. Selain itu, Ruby Chandra juga menggunakan algoritma *Naive Bayes* dan mencapai akurasi 91,64% dengan *precision* sebesar 90,41% dan *recall* sebesar 74,973% [13]. Hasil-hasil tersebut menunjukkan bahwa berbagai algoritma mampu memberikan kinerja yang baik, meskipun tingkat akurasi yang diperoleh berbeda-beda.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengisi *gap* dengan membandingkan kinerja tiga algoritma klasifikasi, yaitu *Naive Bayes*, *Decision Tree*, dan *Random Forest* untuk menentukan metode mana yang memberikan akurasi performa terbaik dalam mengklasifikasikan sentimen pengguna aplikasi Samsat Digital Nasional. Dengan demikian, penelitian ini dapat menjadi dasar bagi pengembang dalam memperbaiki kualitas layanan dan menyempurnakan fitur aplikasi.

## 2. METODE PENELITIAN

Metode penelitian yang digunakan dengan memanfaatkan *Google Colab* dan pustaka pendukung yang tersedia di bahasa pemrograman *Python*. Proses analisis dilakukan dengan pendekatan algoritma *Naive Bayes*, *Decision Tree*, dan *Random Forest* untuk mengevaluasi serta membandingkan tingkat akurasinya. Gambar 1 merupakan alur pelaksanaan penelitian ini secara menyeluruh.



Gambar 1. Alur Penelitian

## 2.1 Pengumpulan Data

Pengumpulan data dilakukan dengan metode *web scraping* untuk memperoleh ulasan pengguna aplikasi Signal – Samsat Digital Nasional yang tersedia pada *Google Play Store*. Proses *web scraping* dijalankan dengan memanfaatkan layanan *Google Colab* sebagai media komputasi berbasis awan (*cloud-based*) yang mendukung bahasa pemrograman *Python* untuk mengakses dan mengekstrak data komentar pengguna aplikasi. Data yang telah terkumpul disimpan dalam format *CSV* dan kemudian diakses kembali melalui *Google Colab* untuk dilakukan analisis [14].

## 2.2 Filter Data

Filter data dilakukan dalam dua tahapan, yaitu pada data mentah sebelum dan sesudah proses *text processing*. Pada tahap pertama, kolom yang tidak relevan dihapus dan dipilih variabel utama penelitian, yaitu *username*, *rating*, dan *review*, serta dilakukan eliminasi data duplikat. Selanjutnya, pada tahap kedua data hasil *text processing* yang meliputi *cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming* kembali difilter untuk memvalidasi dan memastikan tidak ada data tidak valid yang tersisa sebelum masuk ke tahap analisis utama.

## 2.3 Text Processing

Pra-pemrosesan teks merupakan langkah untuk mengoreksi data yang tidak konsisten atau tidak sesuai dalam dataset. Dalam proses ini, berbagai komponen yang tidak relevan akan dihapus, dan data akan diubah menjadi bentuk yang lebih terstruktur [15]. Proses ini mencakup beberapa tahapan sebagai berikut:

- Cleaning** (Pembersihan): Tahapan untuk pembersihan data yang bertujuan untuk mengeliminasi berbagai elemen. Elemen-elemen ini mencakup simbol, karakter khusus seperti (!@#\$%^&\*()[]:;".',/<>?), angka (0-9), dan emotikon [15].
- Case Folding** (Penyeragaman Huruf): Tahapan untuk mengonversi seluruh teks dalam dataset menjadi satu format huruf yang seragam, biasanya dari huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*). Hal ini penting agar kata yang sama namun memiliki kapitalisasi berbeda tidak dianggap sama [15].
- Normalization** (Normalisasi): Tahapan untuk mengoreksi kesalahan pengetikan (*typo*) dan mengubah kata-kata tidak baku, seperti singkatan atau bahasa gaul, menjadi bentuk kata yang baku [15]. Metode ini memanfaatkan kamus *slang* [16] yang berisi pemetaan dari kata tidak baku ke bentuk kata bakunya.
- Tokenizing** (Tokenisasi): Tahapan untuk memecahkan kalimat menjadi unit-unit yang lebih kecil yang disebut "token" sehingga menjadi kata-kata individual yang berdiri sendiri [15].
- Stopword Removal** (Penghapusan Kata): Tahapan untuk menyaring dan membuang kata-kata umum yang sering muncul namun tidak memiliki kontribusi yang signifikan dan memfokuskan pada kata-kata yang penting dan mewakili isi utama dari kalimat [15].

- f. *Stemming* (Pencarian Kata Dasar): Tahapan untuk mengubah kata menjadi bentuk dasarnya dengan menghilangkan semua imbuhan, baik awalan (*prefix*) maupun akhiran (*suffix*) dari kata tersebut [15].

#### 2.4 Pelabelan Data

Pelabelan data dilakukan untuk mengkategorikan ulasan ke dalam dua kelas sentimen, yaitu positif atau negatif, berdasarkan jumlah bobot kata yang dihitung menggunakan kamus *Lexicon InSet* [17]. Penentuan label sentimen didasarkan pada dominasi bobot kata, di mana ulasan diberi label positif apabila bobot kata positif lebih dominan, dan diberi label negatif apabila bobot kata negatif lebih dominan.

#### 2.5 Split Data

Split data dilakukan untuk membagi dataset menjadi dua bagian, yaitu data training (data latih) dan data testing (data uji). Rasio pembagian yang digunakan adalah 80:20, di mana 80% data dialokasikan sebagai data latih untuk melatih model, sedangkan 20% sisanya digunakan sebagai data uji untuk mengevaluasi kinerja objektif [18]. Dataset yang digunakan merupakan hasil pengumpulan data yang telah melalui proses *filtering* sehingga diperoleh 1.743 data bersih. Pembagian tersebut menghasilkan 1.394 data latih dan 349 data uji. Rasio 80:20 dipilih untuk memastikan ketersediaan data yang memadai pada tahap pelatihan serta memungkinkan evaluasi performa model secara objektif.

#### 2.6 Pembobotan TF-IDF

Pembobotan TF-IDF dilakukan untuk menilai bobot sebuah kata dalam dokumen. Metode ini terdiri dari dua matriks perhitungan. Pertama, *Term Frequency* (TF), yaitu kalkulasi frekuensi kemunculan sebuah kata dalam dokumen. Kedua, *Inverse Document Frequency* (IDF), yang berfungsi mengidentifikasi seberapa langka sebuah kata pada seluruh dokumen (korpus) [19]. Adapun rumus perhitungan TF-IDF sebagai berikut:

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

Keterangan:

$W_{dt}$  = nilai bobot dari kata (*term*) ke-t di dalam dokumen ke-d.

$tf_{dt}$  = jumlah kemunculan kata ke-t pada dokumen ke-d.

$IDF_t$  = perbandingan antara jumlah dokumen yang mengandung kata ke-t.

#### 2.7 Klasifikasi Model

Klasifikasi model dilakukan untuk mengevaluasi seberapa baik kinerja model dalam memprediksi data. Proses pemodelan menggunakan tiga algoritma klasifikasi, yaitu *Naïve Bayes*, *Decision Tree*, dan *Random Forest*.

##### a. *Naïve Bayes*

*Naïve Bayes* adalah sebuah algoritma klasifikasi yang bekerja berdasarkan probabilitas. Algoritma ini menghitung peluang suatu data untuk menjadi bagian dari setiap kelas, kemudian memilih kelas dengan peluang terbesar sebagai hasil prediksinya [20]. Adapun rumus *Naïve Bayes* sebagai berikut:

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \quad (2)$$

Keterangan:

$P(X|Y)$  = probabilitas kejadian X terjadi jika diketahui bahwa Y terjadi.

$P(X)$  = probabilitas awal dari kejadian X tanpa mempertimbangkan kondisi.

$P(Y|X)$  = probabilitas kejadian Y terjadi dengan syarat X sudah terjadi.

$P(Y)$  = probabilitas total dari kejadian Y dalam keseluruhan data.

##### b. *Decision Tree*

*Decision Tree* merupakan algoritma klasifikasi yang memodelkan data dalam bentuk pohon keputusan dengan membagi dataset ke dalam cabang-cabang berdasarkan atribut tertentu. Setiap percabangan mewakili keputusan yang diambil menggunakan ukuran seperti *information gain* atau *gini index* untuk menentukan pemisahan terbaik [20]. Adapun rumus *Decision Tree* sebagai berikut:

- f. *Stemming* (Pencarian Kata Dasar): Tahapan untuk mengubah kata menjadi bentuk dasarnya dengan menghilangkan semua imbuhan, baik awalan (*prefix*) maupun akhiran (*suffix*) dari kata tersebut [15].

#### 2.4 Pelabelan Data

Pelabelan data dilakukan untuk mengkategorikan ulasan ke dalam dua kelas sentimen, yaitu positif atau negatif, berdasarkan jumlah bobot kata yang dihitung menggunakan kamus *Lexicon InSet* [17]. Penentuan label sentimen didasarkan pada dominasi bobot kata, di mana ulasan diberi label positif apabila bobot kata positif lebih dominan, dan diberi label negatif apabila bobot kata negatif lebih dominan.

#### 2.5 Split Data

Split data dilakukan untuk membagi dataset menjadi dua bagian, yaitu data training (data latih) dan data testing (data uji). Rasio pembagian yang digunakan adalah 80:20, di mana 80% data dialokasikan sebagai data latih untuk melatih model, sedangkan 20% sisanya digunakan sebagai data uji untuk mengevaluasi kinerja objektif [18]. Dataset yang digunakan merupakan hasil pengumpulan data yang telah melalui proses *filtering* sehingga diperoleh 1.743 data bersih. Pembagian tersebut menghasilkan 1.394 data latih dan 349 data uji. Rasio 80:20 dipilih untuk memastikan ketersediaan data yang memadai pada tahap pelatihan serta memungkinkan evaluasi performa model secara objektif.

#### 2.6 Pembobotan TF-IDF

Pembobotan TF-IDF dilakukan untuk menilai bobot sebuah kata dalam dokumen. Metode ini terdiri dari dua matriks perhitungan. Pertama, *Term Frequency* (TF), yaitu kalkulasi frekuensi kemunculan sebuah kata dalam dokumen. Kedua, *Inverse Document Frequency* (IDF), yang berfungsi mengidentifikasi seberapa langka sebuah kata pada seluruh dokumen (korpus) [19]. Adapun rumus perhitungan TF-IDF sebagai berikut:

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

Keterangan:

$W_{dt}$  = nilai bobot dari kata (*term*) ke-t di dalam dokumen ke-d.

$tf_{dt}$  = jumlah kemunculan kata ke-t pada dokumen ke-d.

$IDF_t$  = perbandingan antara jumlah dokumen yang mengandung kata ke-t.

#### 2.7 Klasifikasi Model

Klasifikasi model dilakukan untuk mengevaluasi seberapa baik kinerja model dalam memprediksi data. Proses pemodelan menggunakan tiga algoritma klasifikasi, yaitu *Naive Bayes*, *Decision Tree*, dan *Random Forest*.

##### a. *Naive Bayes*

*Naive Bayes* adalah sebuah algoritma klasifikasi yang bekerja berdasarkan probabilitas. Algoritma ini menghitung peluang suatu data untuk menjadi bagian dari setiap kelas, kemudian memilih kelas dengan peluang terbesar sebagai hasil prediksinya [20]. Adapun rumus *Naive Bayes* sebagai berikut:

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \quad (2)$$

Keterangan:

$P(X|Y)$  = probabilitas kejadian X terjadi jika diketahui bahwa Y terjadi.

$P(X)$  = probabilitas awal dari kejadian X tanpa mempertimbangkan kondisi.

$P(Y|X)$  = probabilitas kejadian Y terjadi dengan syarat X sudah terjadi.

$P(Y)$  = probabilitas total dari kejadian Y dalam keseluruhan data.

##### b. *Decision Tree*

*Decision Tree* merupakan algoritma klasifikasi yang memodelkan data dalam bentuk pohon keputusan dengan membagi dataset ke dalam cabang-cabang berdasarkan atribut tertentu. Setiap percabangan mewakili keputusan yang diambil menggunakan ukuran seperti *information gain* atau *gini index* untuk menentukan pemisahan terbaik [20]. Adapun rumus *Decision Tree* sebagai berikut:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \times Entropy(S_v) \quad (3)$$

Keterangan:

$Gain(S, A)$  = information gain atribut  $A$  pada dataset  $S$ .

$Entropy(S)$  = ukuran ketidakpastian dalam dataset  $S$ .

$S_v$  = subset data dari  $S$  untuk nilai atribut  $A = v$ .

$|S_v| / |S|$  = proporsi jumlah data pada subset  $S_v$  dibanding total data.

### c. Random Forest

*Random Forest* merupakan metode klasifikasi yang memanfaatkan prinsip *ensemble learning* dengan membangun banyak pohon keputusan. Algoritma ini secara acak memilih subset data dan fitur untuk melatih setiap pohon. Hasil dari semua pohon individu tersebut kemudian *diagregasi* (digabungkan) melalui proses *majority voting* [20]. Adapun rumus *Random Forest* sebagai berikut:

$$H(x) = \text{majority\_vote} \{ h_1(x), h_2(x), \dots, h_k(x) \} \quad (4)$$

Keterangan:

$H(x)$  = hasil prediksi akhir untuk data  $x$ .

$h_1(x)$  = prediksi dari pohon keputusan ke- $i$ .

$k$  = jumlah pohon.

*majority\_vote* = penggabungan hasil prediksi dengan memilih kelas terbanyak

## 2.8 Evaluasi Algoritma

Evaluasi algoritma dilakukan untuk menilai kinerja model klasifikasi yang telah dibangun. Proses ini memanfaatkan metode *Confusion Matrix*, yang merupakan sebuah tabel untuk membandingkan hasil prediksi model dengan label data yang aktual [21]. Adapun rumus *Confusion Matrix* sebagai berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (8)$$

Keterangan:

TP: Prediksi positif yang tepat untuk data aktual positif.

TN: Prediksi negatif yang tepat untuk data aktual negatif.

FP: Prediksi positif yang keliru untuk data aktual negatif.

FN: Prediksi negatif yang keliru untuk data aktual positif

## 2.9 Visualisasi Data

Visualisasi data dilakukan untuk memudahkan pemahaman pola dalam penelitian. Visualisasi ini meliputi distribusi sentimen yang menampilkan kategori sentimen (positif atau negatif), distribusi rating yang menampilkan tingkat kepuasan pengguna, serta *word cloud* yang menampilkan kata-kata paling dominan [22].

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Pada tahap ini, peneliti menggunakan metode *web scraping* untuk memperoleh ulasan pengguna aplikasi Signal – Samsat Digital Nasional yang tersedia pada *Google Play Store*. Jumlah data yang di *scraping* sebanyak 2000 ulasan yang kemudian digunakan sebagai dataset dalam penelitian.



**Gambar 2.** Alur Pengumpulan Data

Gambar 2 menampilkan alur proses pengumpulan data menggunakan metode *web scraping* pada ulasan pengguna aplikasi SAMSAT di Google Play Store. Data dikumpulkan secara otomatis menggunakan *script Python* di *Google Colab* dan disimpan dalam *format CSV*. Dataset tersebut kemudian diproses menggunakan *script Python* di *Google Colab* untuk tahap pengolahan data. Hasil pengumpulan data ditampilkan pada Tabel 1.

**Tabel 1.** Ilustrasi Hasil *Web Scraping*

| <i>index</i> | <i>reviewId</i>              | <i>userName</i>   | <i>userImage</i>                          | <i>content</i>                                | <i>score</i> | <i>thumbsUp Count</i> | <i>version</i> |
|--------------|------------------------------|-------------------|-------------------------------------------|-----------------------------------------------|--------------|-----------------------|----------------|
| 1            | cd33c99e-1a2a-4aa5-....      | Ridwan Firmansyah | https://play-lh.googleusercontent.com/a-/ | Aplikasi signal ini sangat membantu sekali... | 5            | 0                     | 1.7.3          |
| 2            | 145f0f6d-b7d1-4c8b-b4cd-.... | Ria Apriana       | https://play-lh.googleusercontent.com/a-/ | ini aplikasi pemerintah kok gini sih?...      | 1            | 0                     | 1.7.3          |

Tabel 1 menampilkan hasil pengumpulan ulasan pengguna yang berisi informasi seperti *reviewId*, *userName*, *userImage*, *content*, *score*, *thumbsUpCount* dan *version* yang menggambarkan identitas, isi, serta penilaian pengguna terhadap aplikasi. Data yang diperoleh masih bersifat mentah (*raw*) dan akan digunakan sebagai dasar untuk tahap pengolahan dan analisis selanjutnya.

### 3.2 Filter Data

Pada tahap ini, data hasil *web scraping* yang masih mentah (*raw*) difilter agar bersih. Proses ini mencakup penghapusan duplikat, entri kosong, serta penyortiran data. Kolom utama yang digunakan adalah *username*, *rating*, dan *review*, sedangkan kolom lain seperti *reviewId*, *userImage*, *thumbsUpCount*, dan *version* seperti pada Tabel 1 diabaikan karena tidak berpengaruh terhadap analisis. Hasil pemilihan kolom ditampilkan pada Tabel 2.

**Tabel 2.** Ilustrasi Filter Pemilihan Kolom

| <i>username</i>   | <i>rating</i> | <i>review</i>                                                                                                 |
|-------------------|---------------|---------------------------------------------------------------------------------------------------------------|
| Ridwan Firmansyah | 5             | Aplikasi signal ini sangat membantu sekali apalagi kami yang ada di daerah kabupaten yang belum ada samsatnya |
| Ria Apriana       | 1             | ini aplikasi pemerintah kok gini sih? mau daftar pengesahan STNK malah muncul notif tidak berhasil            |

Tabel 2 menampilkan hasil data ulasan pengguna setelah melalui tahap *filtering*. Pada tahap ini, hanya kolom *username*, *rating*, dan *review* yang dipertahankan sebagai variabel utama penelitian. Sementara itu, kolom lain seperti *reviewId*, *userImage*, *thumbsUpCount*, dan *version* dihapus. Adapun total data setelah melakukan filter data ditampilkan pada Gambar 3.

| Hasil Pengumpulan Data                                                                                                                                                                                                                                                                                                                           | Filter Sebelum Text-Processing                                                                                                                                                                                                                                           | Filter Sesudah Text-Processing                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre>Data columns (total 5 columns): #  Column      Non-Null Count  Dtype ---  - 0  id            2000 non-null   object 1  username      2000 non-null   object 2  rating        2000 non-null   int64 3  review        2000 non-null   object 4  date          2000 non-null   object dtypes: int64(1), object(4) memory usage: 78.3+ KB</pre> | <pre>Data columns (total 3 columns): #  Column      Non-Null count  Dtype ---  - 0  username     1753 non-null   object 1  rating       1753 non-null   int64 2  review       1753 non-null   string dtypes: int64(1), object(1), string(1) memory usage: 41.2+ KB</pre> | <pre>Data columns (total 9 columns): #  Column      Non-Null Count  Dtype ---  - 0  username     1743 non-null   object 1  rating       1743 non-null   int64 2  review       1743 non-null   string 3  cleaning     1743 non-null   object 4  case_folding 1743 non-null   object 5  normalization 1743 non-null   object 6  tokenize     1743 non-null   object 7  stopwords_removal 1743 non-null object 8  stemming     1743 non-null   object dtypes: int64(1), object(7), string(1) memory usage: 115.9+ KB</pre> |

**Gambar 3. Hasil Penghapusan Data Duplikat/Kosong**

Gambar 3 menampilkan hasil proses filter data yang mencakup penghapusan data duplikat dan data kosong. Pada tahap awal, diperoleh 2.000 data dengan lima kolom utama, kemudian difilter menjadi 1.753 data dengan tiga kolom, yaitu *username*, *rating*, dan *review*. Selanjutnya, setelah tahap text processing yang meliputi *cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming*, sehingga diperoleh 1.743 data bersih siap digunakan untuk analisis.

### 3.3 Text Processing

Pada tahap ini, data hasil *filtering* dilanjutkan ke *text processing* untuk mengubah data tidak terstruktur menjadi lebih terstruktur. Proses ini dilakukan untuk menormalkan data sehingga dapat diolah oleh algoritma *machine learning*. Beberapa langkah yang dilakukan adalah sebagai berikut:

**Tabel 3. Ilustrasi Hasil Text Processing**

| Tahapan                 | Ulasan                                                                                                                                                                                                |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Data Awal</b>        | Sya udah bayar pajak motor lewat aplikasi ini tapi udah hampir 3 bulan blm sampe ke rumah....gimana ini.....kok malah jadi ribet                                                                      |
| <b>Cleaning</b>         | Sya udah bayar pajak motor lewat aplikasi ini tapi udah hampir bulan blm sampe ke rumah gimana ini kok malah jadi ribet                                                                               |
| <b>Case Folding</b>     | sya udah bayar pajak motor lewat aplikasi ini tapi udah hampir bulan blm sampe ke rumah gmn ini kok malah jadi ribet                                                                                  |
| <b>Normalization</b>    | saya sudah bayar pajak motor lewat aplikasi ini tetapi sudah hampir bulan belum sampai ke rumah bagaimana ini kok malah jadi ribet                                                                    |
| <b>Tokenizing</b>       | ['saya', 'sudah', 'bayar', 'pajak', 'motor', 'lewat', 'aplikasi', 'ini', 'tetapi', 'sudah', 'hampir', 'bulan', 'belum', 'sampai', 'ke', 'rumah', 'bagaimana', 'ini', 'kok', 'malah', 'jadi', 'ribet'] |
| <b>Stopword Removal</b> | ['bayar', 'pajak', 'motor', 'aplikasi', 'rumah', 'ribet']                                                                                                                                             |
| <b>Stemming</b>         | bayar pajak motor aplikasi rumah ribet                                                                                                                                                                |

Tabel 3 menampilkan tahapan *text processing* pada data ulasan, mulai dari *cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal* dan *stemming*. Setiap tahap ditunjukkan secara berurutan sehingga teks mentah berubah menjadi data yang siap untuk analisis.

### 3.4 Pelabelan Data

Pada tahap ini, pelabelan data dilakukan untuk mengkategorikan kelas sentimen sebagai dasar penelitian. Proses ini menggunakan kamus *Lexicon InSet* [20] untuk memberikan bobot pada setiap kata sehingga memudahkan proses pengkategorian secara otomatis. Ulasan diberi label positif jika bobot kata positif lebih dominan dan negatif jika bobot kata negatif lebih dominan. Hasil pelabelan data ditampilkan pada Tabel 4.

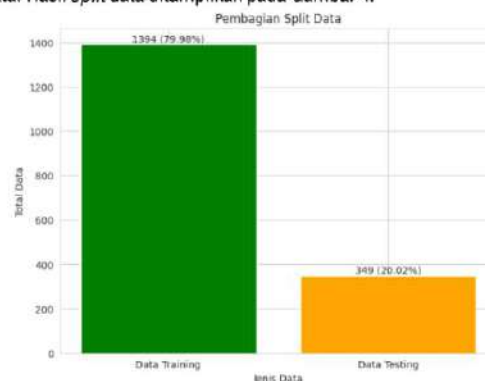
**Tabel 4. Ilustrasi Hasil Transformasi Pelabelan Data**

| Hasil Stemming                                        | Sentimen |
|-------------------------------------------------------|----------|
| gagal bayar pajak daftar gagal tolong sistem ganggu   | Negatif  |
| bayar aplikasi transfer apa tolong tindak lanjut      |          |
| aplikasi bantu bayar pajak tahun kendaru daring mudah | Positif  |
| mana kota daerah moga depan mudah lancar              |          |

Tabel 4 menampilkan hasil *stemming* terhadap ulasan pengguna yang telah diberi label sentimen menggunakan *Lexicon InSet*. Ulasan dengan bobot kata yang mengandung keluhan atau pengalaman kurang menyenangkan dikategorikan sebagai sentimen negatif, sedangkan ulasan dengan bobot kata yang menunjukkan apresiasi atau pujian dikategorikan sebagai sentimen positif.

### 3.5 Split Data

Pada tahap ini, dataset dibagi menjadi dua bagian, yaitu data *training* (alokasi 80%) dan data *testing* (alokasi 20%) melalui proses *split* data. Hasil *split* data ditampilkan pada Gambar 4.



Gambar 4. Hasil Pembagian Split Data

Gambar 4 menampilkan hasil pembagian dataset menjadi 1.394 data latih (80%) dan 349 data uji (20%). Rasio 80:20 dipilih agar sebagian besar data dapat digunakan untuk proses pelatihan model, sementara sisanya dimanfaatkan untuk menguji dan mengevaluasi performa model secara objektif.

### 3.6 Pembobotan TF-IDF

Pada tahap ini, dilakukan pembobotan TF-IDF untuk menilai tingkat kepentingan setiap kata dalam dokumen. Metode ini menghitung bobot kata berdasarkan frekuensi kemunculan (*Term Frequency*) dan tingkat kekhususannya dalam seluruh dokumen (*Inverse Document Frequency*), sehingga dapat mengidentifikasi kata-kata yang paling relevan dan berpengaruh terhadap analisis. Hasil transformasi pembobotan TF-IDF ditampilkan pada Tabel 5.

Tabel 5. Ilustrasi Hasil TF-IDF

| Index | acab | account | app | adalagi | apdet | calo | ... | jansip | admin |
|-------|------|---------|-----|---------|-------|------|-----|--------|-------|
| 0     | 0.0  | 0.0     | 0.0 | 0.0     | 0.0   | 0.0  | ... | 0.0    | 0.0   |
| 1     | 0.0  | 0.0     | 0.0 | 0.0     | 0.0   | 0.0  | ... | 0.0    | 0.0   |
| 2     | 0.0  | 0.0     | 0.0 | 0.0     | 0.0   | 0.0  | ... | 0.0    | 0.0   |
| 3     | 0.0  | 0.0     | 0.0 | 0.0     | 0.0   | 0.0  | ... | 0.0    | 0.0   |
| 4     | 0.0  | 0.0     | 0.0 | 0.0     | 0.0   | 0.0  | ... | 0.0    | 0.0   |
| 5     | 0.0  | 0.0     | 0.0 | 0.0     | 0.0   | 0.0  | ... | 0.0    | 0.0   |

Tabel 5 menampilkan hasil pembobotan TF-IDF, yang menunjukkan bobot setiap kata berdasarkan seberapa sering kata itu muncul. Apabila sebuah kata memiliki nilai bobot 0, ini mengindikasikan bahwa kata tersebut tidak ditemukan atau tidak memberi kontribusi pada dokumen.

### 3.7 Klasifikasi Model

Pada tahap ini, dilakukan klasifikasi untuk mengevaluasi kinerja model dalam memprediksi data berdasarkan pembobotan TF-IDF. Proses pemodelan menggunakan tiga algoritma, yaitu *Naive Bayes*, *Decision Tree*, dan *Random Forest*. Hasil klasifikasi model ditampilkan pada Gambar 5.

```

from sklearn.naive_bayes import MultinomialNB
# Latih Model Naive Bayes Dengan Data Training
nb = MultinomialNB()
nb.fit(tfidf_training, y_training)
# Prediksi Data Testing
y_pred = nb.predict(tfidf_testing)

- MultinomialNB
MultinomialNB()

from sklearn.tree import DecisionTreeClassifier
# Latih Model Decision Tree Dengan Data Training
tree = DecisionTreeClassifier(random_state=42)
tree.fit(tfidf_training, y_training)
# Prediksi Data Testing
y_pred_tree = tree.predict(tfidf_testing)

- DecisionTreeClassifier
DecisionTreeClassifier()

from sklearn.ensemble import RandomForestClassifier
# Latih Model Random Forest Dengan Data Training
forest = RandomForestClassifier(n_estimators=100, random_state=42)
forest.fit(tfidf_training, y_training)
# Prediksi Data Testing
y_pred_forest = forest.predict(tfidf_testing)

- RandomForestClassifier
RandomForestClassifier(random_state=42)

```

Gambar 5. Modelling Algoritma

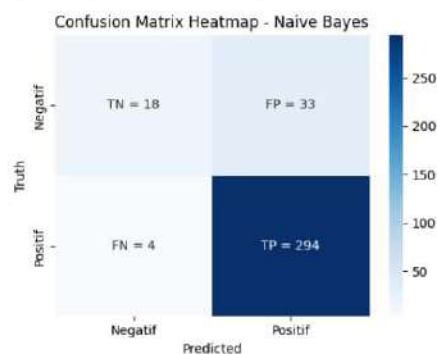
Gambar 5 menampilkan proses modelling menggunakan tiga algoritma klasifikasi dari *library scikit-learn*, yaitu *Naïve Bayes* dengan parameter *MultinomialNB()*, *Decision Tree* dengan parameter *DecisionTreeClassifier()*, dan *Random Forest* dengan parameter *RandomForestClassifier()*. Model dilatih menggunakan perintah *fit(tfidf\_training, y\_training)* pada data *training* dan diuji dengan perintah *predict(tfidf\_testing)* pada data *testing*. Hasil dari ketiga model ini nantinya dibandingkan untuk mengevaluasi performa dan akurasi algoritma terbaik.

### 3.8 Evaluasi Algoritma

Pada tahap ini, dilakukan pengukuran kinerja terhadap model klasifikasi yang sudah dibangun. Model-model ini menggunakan tiga algoritma berbeda: *Naïve Bayes*, *Decision Tree*, dan *Random Forest*. Rincian dari hasil evaluasi ketiga algoritma itu adalah sebagai berikut.

#### 3.8.1 Evaluasi Algoritma *Naïve Bayes*

Evaluasi algoritma *Naïve Bayes* divisualisasikan melalui *confusion matrix* pada Gambar 6, yang menunjukkan kemampuan model dalam memprediksi kelas positif dan negatif. Berdasarkan hasil tersebut, model menghasilkan 294 *True Positive* (TP), 4 *False Negative* (FN), 18 *True Negative* (TN), dan 33 *False Positive* (FP).



Gambar 6. Hasil Evaluasi *Naïve Bayes*

Gambar 6 menampilkan *confusion matrix heatmap* dari algoritma *Naïve Bayes* yang menunjukkan bahwa model memiliki kemampuan lebih baik dalam memprediksi kelas positif dibandingkan kelas negatif. Adapun hasil perhitungan evaluasi setiap kelas disajikan sebagai berikut:

#### 1. Evaluasi Kelas Positif

$$Precision = \frac{TP}{TP + FP} = \frac{294}{294 + 33} = \frac{294}{327} = 0.899 \text{ (89,9\%)}$$

$$Recall = \frac{TP}{TP + FN} = \frac{294}{294 + 4} = \frac{294}{298} = 0.987 \text{ (98.7\%)}$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{(0.899 \times 0.987)}{0.899 + 0.987} = 0.939 (93.9\%)$$

## 2. Evaluasi Kelas Negatif

$$Precision = \frac{TN}{TN + FN} = \frac{18}{18 + 4} = \frac{18}{22} = 0.818 (81.8\%)$$

$$Recall = \frac{TN}{TN + FP} = \frac{18}{18 + 33} = \frac{18}{51} = 0.353 (35.3\%)$$

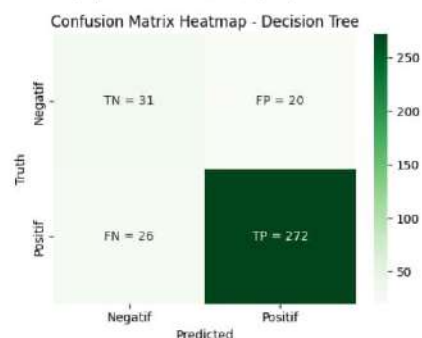
$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{(0.818 \times 0.353)}{0.818 + 0.353} = 0.491 (49.1\%)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{294 + 18}{294 + 18 + 33 + 4} = \frac{312}{349} = 0.894 (89.4\%)$$

Dengan akurasi keseluruhan 89,4%, performa *Naïve Bayes* terbilang cukup baik. Model ini menunjukkan keandalan tinggi dalam mendeteksi sentimen positif, yang dibuktikan dengan nilai *precision* 89,9%, *recall* 98,7%, dan *f1-score* 93,9%. Akan tetapi, kinerjanya menurun drastis saat mendeteksi sentimen negatif, di mana nilai *recall*-nya hanya 35,3% (dengan *precision* 81,8% dan *f1-score* 49,1%). Dapat disimpulkan bahwa algoritma ini sangat baik untuk prediksi positif, tetapi lemah untuk prediksi negatif.

### 3.8.2 Evaluasi Algoritma Decision Tree

Evaluasi algoritma Decision Tree divisualisasikan melalui confusion matrix pada Gambar 7, yang menunjukkan kemampuan model dalam memprediksi kelas positif dan negatif. Berdasarkan hasil tersebut, model menghasilkan 272 True Positive (TP), 26 False Negative (FN), 31 True Negative (TN), dan 20 False Positive (FP).



Gambar 7. Hasil Evaluasi Decision Tree

Gambar 7 menampilkan *confusion matrix heatmap* dari algoritma *Decision Tree* yang menunjukkan bahwa model memiliki kemampuan lebih baik dalam memprediksi kelas positif dibandingkan kelas negatif. Adapun hasil perhitungan evaluasi setiap kelas disajikan sebagai berikut:

## 1. Evaluasi Kelas Positif

$$Precision = \frac{TP}{TP + FP} = \frac{272}{272 + 20} = \frac{272}{292} = 0.932 (93.2\%)$$

$$Recall = \frac{TP}{TP + FN} = \frac{272}{272 + 26} = \frac{272}{298} = 0.913 (91.3\%)$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{(0.932 \times 0.913)}{0.932 + 0.913} = 0.922 (92.2\%)$$

## 2. Evaluasi Kelas Negatif

$$Precision = \frac{TN}{TN + FN} = \frac{31}{31 + 26} = \frac{31}{57} = 0.544 (54.4\%)$$

$$Recall = \frac{TN}{TN + FP} = \frac{31}{31 + 20} = \frac{31}{51} = 0.608 (60.8\%)$$

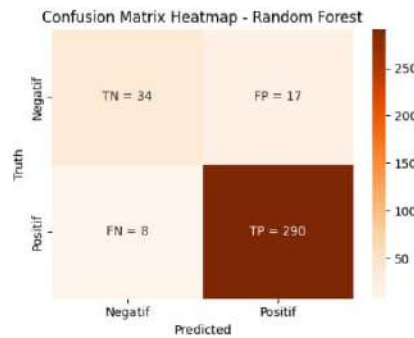
$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{(0.544 \times 0.608)}{0.544 + 0.608} = 0.574 (57.5\%)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{272 + 31}{272 + 31 + 20 + 26} = \frac{303}{349} = 0.868 (86.8\%)$$

Dengan akurasi keseluruhan 86,8%, performa *Decision Tree* terbilang cukup baik. Model ini menunjukkan keandalan tinggi dalam mendeteksi sentimen positif, yang dibuktikan dengan nilai *precision* 93,2%, *recall* 91,3%, dan *f1-score* 92,2%. Akan tetapi, kinerjanya menurun saat mendeteksi sentimen negatif, di mana nilai *precision* hanya 54,4% (dengan *recall* 60,8% dan *f1-score* 57,5%). Dapat disimpulkan bahwa algoritma ini sangat baik untuk prediksi positif, namun masih kurang konsisten dalam mengidentifikasi sentimen negatif.

### 3.8.3 Evaluasi Algoritma Random Forest

Evaluasi algoritma *Random Forest* divisualisasikan melalui confusion matrix pada Gambar 8, yang menunjukkan kemampuan model dalam memprediksi kelas positif dan negatif. Berdasarkan hasil tersebut, model menghasilkan 290 *True Positive* (TP), 8 *False Negative* (FN), 34 *True Negative* (TN), dan 17 *False Positive* (FP).



Gambar 8. Hasil Evaluasi Random Forest

Gambar 8 menampilkan *confusion matrix heatmap* dari algoritma *Random Forest* yang menunjukkan bahwa model memiliki kemampuan lebih baik dalam memprediksi kelas positif dibandingkan kelas negatif. Adapun hasil perhitungan evaluasi setiap kelas disajikan sebagai berikut:

## 1. Evaluasi Kelas Positif

$$Precision = \frac{TP}{TP + FP} = \frac{290}{290 + 17} = \frac{290}{307} = 0.945 (94.5\%)$$

$$Recall = \frac{TP}{TP + FN} = \frac{290}{290 + 8} = \frac{290}{298} = 0.973 (97.3\%)$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{(0.945 \times 0.973)}{0.945 + 0.973} = 0.959 (95.9\%)$$

## 2. Evaluasi Kelas Negatif

$$Precision = \frac{TN}{TN + FN} = \frac{34}{34 + 8} = \frac{34}{42} = 0.810 (81\%)$$

$$Recall = \frac{TN}{TN + FP} = \frac{34}{34 + 17} = \frac{34}{51} = 0.667 (66.7\%)$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} = 2 \times \frac{(0.810 \times 0.667)}{0.810 + 0.667} = 0.731 (73.1\%)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{290 + 34}{290 + 34 + 17 + 8} = \frac{324}{349} = 0.929 (92.9\%)$$

Dengan akurasi keseluruhan 92,9%, performa *Random Forest* terbukti menjadi yang terbaik dibandingkan algoritma lainnya. Model ini menunjukkan keandalan sangat tinggi dalam mendeteksi sentimen positif, yang dibuktikan dengan nilai *precision* 94,5%, *recall* 97,3%, dan *f1-score* 95,9%. Selain itu, kinerjanya saat mendeteksi sentimen negatif juga jauh lebih baik, dengan *precision* 81,0%, *recall* 66,7%, dan *f1-score* 73,1%. Dapat disimpulkan bahwa algoritma ini memiliki kemampuan terbaik dalam mengklasifikasikan sentimen negatif di antara algoritma yang diuji.

### 3.9 Hasil Perbandingan Algoritma

Evaluasi terhadap algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest* menunjukkan perbedaan performa dalam mengklasifikasikan sentimen. Setiap model memiliki efektivitas yang berbeda dalam membedakan sentimen positif dan negatif, dengan rincian perbandingan akurasi yang ditampilkan pada Tabel 6.

Tabel 6. Hasil Perbandingan Algoritma

| Metrik           | Algoritma          |         |                      |         |                      |         |
|------------------|--------------------|---------|----------------------|---------|----------------------|---------|
|                  | <i>Naïve Bayes</i> |         | <i>Decision Tree</i> |         | <i>Random Forest</i> |         |
|                  | Sentimen           |         | Sentimen             |         | Sentimen             |         |
|                  | Positif            | Negatif | Positif              | Negatif | Positif              | Negatif |
| <b>Precision</b> | 0.899              | 0.820   | 0.932                | 0.544   | 0.945                | 0.810   |
| <b>Recall</b>    | 0.987              | 0.353   | 0.913                | 0.608   | 0.973                | 0.667   |
| <b>F1-score</b>  | 0.941              | 0.491   | 0.922                | 0.574   | 0.959                | 0.731   |
| <b>Accuracy</b>  | <b>0.894</b>       |         | <b>0.868</b>         |         | <b>0.929</b>         |         |

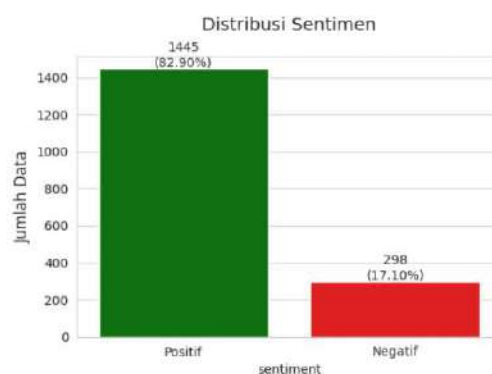
Tabel 6 menampilkan perbandingan hasil evaluasi tiga algoritma klasifikasi. *Random Forest* menunjukkan performa terbaik dengan akurasi tertinggi sebesar 92,9%, mengungguli *Naïve Bayes* (89,4%) dan *Decision Tree* (86,8%). Algoritma *Random Forest* lebih seimbang dalam mengenali sentimen positif dan negatif, sehingga menjadi yang paling efektif dalam penelitian ini.

### 3.10 Visualisasi Data

Pada tahap ini, visualisasi data ditampilkan untuk memudahkan pemahaman pola yang terkandung dalam dataset. Dibawah ini ditampilkan visualisasi yang meliputi distribusi sentimen, distribusi rating dan *word cloud* yang menampilkan kata-kata paling dominan yang dijabarkan sebagai berikut.

#### 3.10.1 Distribusi Sentimen

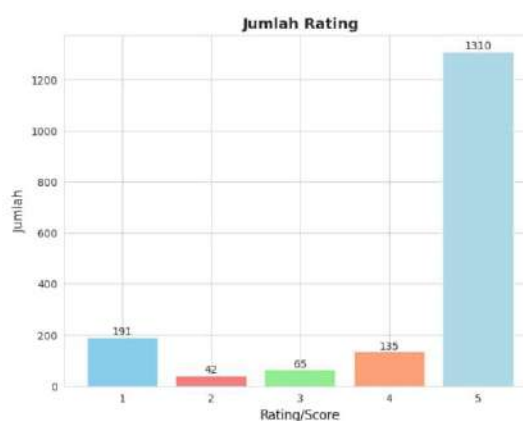
Distribusi sentimen menunjukkan ketidakseimbangan data, dari total 1743 data di mana sentimen positif (1.445 data atau 82,90%) jauh lebih dominan dibandingkan sentimen negatif (298 data atau 17,10%). Klasifikasi ini dilakukan menggunakan kamus *lexicon*. Hasil distribusi sentimen tersebut ditampilkan pada Gambar 9.



Gambar 9. Hasil Distribusi Sentimen

### 3.10.2 Distribusi Rating

Distribusi rating menampilkan rating dari 1 hingga 5, beserta jumlah pengguna yang memberikan ulasan. Dari total 1.743 data, terdapat rating 1 (191 data), rating 4 (135 data), rating 3 (65 data), dan rating 2 (42 data) dan rating 5 (1.310 data). Hasil distribusi rating tersebut ditampilkan pada Gambar 10.



Gambar 10. Hasil Distribusi Rating

### 3.10.3 Word Cloud

Dalam wordcloud ini, terlihat kata-kata yang sering digunakan oleh pengguna saat memberikan ulasan, seperti kata "bantu," "pajak," "mudah," dan "cepat. Kata-kata ini mencerminkan aspek yang dirasakan pengguna, seperti kemudahan, kenyamanan serta kepuasan terhadap aplikasi. Hasil visualisasi *word cloud* ditampilkan pada Gambar 11.



#### 4. KESIMPULAN

## DAFTAR PUSTAKA

- [1] M. S. J. Sangaji and J. Irianto, "Transformasi Inovasi Pelayanan Publik menuju Pemerintahan Digital," *Jejaring Administrasi Publik*, vol. 17, no. 1, pp. 54–70, Jun. 2025, doi: 10.20473/jap.v17i1.72708.
- [2] S. N. Ufa and A. Frinaldi, "Inovasi Pemerintah Daerah Dalam Penerapan Kebijakan Publik Dan Pelayanan Publik," *SOCIAL : Jurnal Inovasi Pendidikan IPS*, vol. 5, no. 2, pp. 610–621, Jun. 2025, doi: 10.51878/social.v5i2.5965.
- [3] D. Wijaya, R. A. Saputra, and F. Irwiensyah, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 4, pp. 2369–2380, Feb. 2024, doi: 10.30865/klk.v4i4.1738.
- [4] A. Afifudin, A. Kadir, and I. Isnaini, "Implementasi Kebijakan Sistem Administrasi Manunggal Satu Atap di Kabupaten Mandailing Natal," *Strukturasi: Jurnal Ilmiah Magister Administrasi Publik*, vol. 2, no. 1, pp. 59–69, Apr. 2020, doi: 10.31289/strukturasi.v2i1.43.
- [5] Suriyani BB, La Ode Agus Said, and Astriana Syaqira, "Digitalisasi Layanan Samsat Peningkatan Kepatuhan Pajak Kendaraan Bermotor Melalui e-Samsat di Kabupaten Muna," *Journal Publicuho*, vol. 7, no. 4, pp. 2111–2126, Dec. 2024, doi: 10.35817/publicuho.v7i4.573.
- [6] Afandi Afandi, Zainal fatah, Sri Roekminiati, and Ika Devy Pramudiana, "Inovasi Layanan Publik dalam Era Digital: Tinjauan Persepsi Generasi Muda pada Samsat Digital di Lumajang," *SOSIAL : Jurnal Ilmiah Pendidikan IPS*, vol. 3, no. 3, pp. 265–289, Jun. 2025, doi: 10.62383/sosial.v3i3.1064.
- [7] C. Triyosanti, D. F. Mufarokhah, S. N. Angelina, A. S. Cahyono, and D. I. Margayaningsih, "Dampak Penerapan Aplikasi Layanan Publik (Samsat Digital Nasional) Terhadap Pelayanan Pajak Kendaraan

- Bermotor di Tulungagung," *JSPH: Jurnal Sosial Politik Humaniora*, vol. 1, no. 3, pp. 42–49, Dec. 2024, doi: 10.59966/jsph.v1i3.1426.
- [8] F. F. Irfani, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Informatika)*, vol. 16, no. 3, pp. 258–266, Feb. 2020, doi: 10.26487/jbmi.v16i3.8607.
- [9] R. Chandra and E. M. Sipayung, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Menggunakan Algoritma Naive Bayes Classifier," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 3, pp. 156–164, Jan. 2025, doi: 10.25077/TEKNOSI.v10i3.2024.156-164.
- [10] A. Musthafa, T. Harmini, A. Rafiq, and N. Marantika, "Pemanfaatan Machine Learning dalam Menganalisis Sentimen Terhadap Program TAPERA di Platform Digital X," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 587–597, Mar. 2025, doi: 10.57152/malcom.v5i2.1801.
- [11] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digital Nasional (Signal) Dengan Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4073.
- [12] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digital Nasional (Signal) Dengan Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4073.
- [13] S. Kacung, C. Pamungkas Putra Bagayana, and D. Cahyono, "Analisis Sentimen Terhadap Layanan Samsat Digital Nasional (Signal) Menggunakan Metode SVM," *Jurnal Mnemonic*, vol. 7, no. 1, pp. 118–122, Apr. 2024, doi: 10.36040/mnemonic.v7i1.9557.
- [14] S. A. R. Rizaldi, S. Alam, and I. Kurniawan, "Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 109–117, Aug. 2023, doi: 10.55123/storage.v2i3.2334.
- [15] Trisnaeni Faradaningsih and Anisa Lutfiyani, "Perbandingan Analisis Sentimen Pengguna Aplikasi Shopee dan Lazada pada Situs Google Play Store Menggunakan Algoritma K-Nearest Neighbor dan Naive Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 4, no. 3, pp. 563–578, Jun. 2025, doi: 10.55123/insologi.v4i3.5646.
- [16] Fornigulo, "Kamus Slang." Accessed: Mar. 05, 2025. [Online]. Available: <https://kaggle.com/datasets/fornigulo/kamus-slag>
- [17] F. Koto and G. Y. Rahmanningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, IEEE, Dec. 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625.
- [18] A. S. Pamungkas and N. Cahyono, "Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 1–10, Jun. 2024, doi: 10.29408/edumatic.v8i1.24114.
- [19] Gilbert, Syariful Alam, and M. Imam Sulistyio, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan Metode Naive Bayes," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 100–108, Aug. 2023, doi: 10.55123/storage.v2i3.2333.
- [20] I. N. Sari, M. K. L. F. Lidimilah, and M. K. A. Lutfi, "Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naive Bayes dalam Klasifikasi Gangguan Tidur," *Proceedings National Conference Sinesia*, vol. 1, no. 1, pp. 192–209, Jun. 2025, doi: 10.69836/ncrcs-sinesia.v1i1.41.
- [21] M. Tirta Nugraha, N. Nina Sulistyowati, and U. Ultach Enri, "Analisis Sentimen Ulasan Aplikasi Satu Sehat Pada Google Play Store Menggunakan Naive Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3593–3601, Jan. 2024, doi: 10.36040/jati.v7i5.7753.
- [22] N. Cahyono and Anggista Oktavia Praneswara, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *Indonesian Journal of Computer Science*, vol. 12, no. 6, Dec. 2023, doi: 10.33022/ijcs.v12i6.3473.