



## **KERJA PRAKTIK**

### **LAPORAN KERJA PRAKTIK**

#### **KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP LAYANAN KPP PRATAMA SURABAYA GUBENG BERBASIS NAIVE BAYES DAN SVM**

#### **KPP PRATAMA SURABAYA GUBENG**

**Oleh:**

|                            |                        |
|----------------------------|------------------------|
| <b>Arifah Shafyan</b>      | <b>NRP. 3322600003</b> |
| <b>Puput Ayu Setiawati</b> | <b>NRP. 3322600004</b> |

**Dosen Pembimbing:**

|  |                                |
|--|--------------------------------|
| <b>Fitrah Maharani Humaira, M.Kom</b>  | <b>NIP. 198906202022032008</b> |
| <b>Ahmad Syauqi Ahsan, S.Kom., MT.</b> | <b>NIP. 197505302003121001</b> |

**PROGRAM STUDI D4 SAINS DATA TERAPAN  
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA  
2025**

Halaman ini sengaja dikosongkan

# **LAPORAN KERJA PRAKTIK**

## **KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP LAYANAN KPP PRATAMA SURABAYA GUBENG BERBASIS NAIVE BAYES DAN SVM**

Periode 13 Januari – 9 Mei 2025



Oleh:

Arifah Shafyan

NRP. 3322600003

Puput Ayu Setiawati

NRP. 3322600004

**PROGRAM STUDI D4 SAINS DATA TERAPAN  
DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER  
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA  
2025**

Halaman ini sengaja dikosongkan

# **HALAMAN PENGESAHAN KERJA PRAKTIK**

## **KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP LAYANAN KPP PRATAMA SURABAYA GUBENG BERBASIS NAIVE BAYES DAN SVM**

KPP PRATAMA SURABAYA GUBENG  
Jl. Sumatera No.22-24, Gubeng, Surabaya, Jawa Timur  
Tanggal Pelaksanaan : 13 Januari 2025 – 09 Mei 2025

### **Oleh:**

Arifah Shafyan                      3322600003

Surabaya, 13 Juni 2025

### **Menyetujui:**

Dosen Pembimbing

Koordinator Kerja Praktek

Ahmad Syauqi Ahsan, S.Kom., MT.  
NIP. 197505302003121001

Alfi Fadliana S.Si., M.Stat  
NIP. 199211252022032018

### **Mengetahui:**

Ketua Program Studi  
D4 Sains Data Terapan

Isbat Uzzin Nadhori S.Kom., M.T  
NIP. 197405052003121002

Halaman ini sengaja dikosongkan

**HALAMAN PENGESAHAN**  
**KERJA PRAKTIK**

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP  
LAYANAN KPP PRATAMA SURABAYA GUBENG BERBASIS  
NAIVE BAYES DAN SVM**

KPP PRATAMA SURABAYA GUBENG  
Jl. Sumatera No.22-24, Gubeng, Surabaya, Jawa Timur  
Tanggal Pelaksanaan : 13 Januari 2025 – 09 Mei 2025

**Oleh:**

Puput Ayu Setiawati                      3322600004

Surabaya, 13 Juni 2025

**Menyetujui:**

Dosen Pembimbing                      Koordinator Kerja Praktek

Fitrah Maharani Humaira, M.Kom  
NIP. 198906202022032008

Alfi Fadliana S.Si., M.Stat  
NIP. 199211252022032018

**Mengetahui:**

Ketua Program Studi  
D4 Sains Data Terapan

Isbat Uzzin Nadhori S.Kom., M.T  
NIP. 197405052003121002

Halaman ini sengaja dikosongkan



**HALAMAN PENGESAHAN  
KERJA PRAKTIK**

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP  
LAYANAN KPP PRATAMA SURABAYA GUBENG BERBASIS  
NAIVE BAYES DAN SVM**

KPP PRATAMA SURABAYA GUBENG  
Jl. Sumatera No.22-24, Gubeng, Surabaya, Jawa Timur  
Tanggal Pelaksanaan : 13 Januari 2025 – 09 Mei 2025

**Oleh:**

Arifah Shafyan                      3322600003

Surabaya, 13 Juni 2025

**Menyetujui:**

Pembimbing Kerja Praktek

**Ahmad Syauqi Ahsan, S.Kom., MT.**

**NIP. 197505302003121001**

**Mengetahui:**

Pembimbing Lapangan  
KPP Pratama Surabaya Gubeng

**Diah Kusuma Sugiarto**

**NIP. 810450527**

Halaman ini sengaja dikosongkan

**HALAMAN PENGESAHAN  
KERJA PRAKTEK**

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP  
LAYANAN KPP PRATAMA SURABAYA GUBENG BERBASIS  
NAIVE BAYES DAN SVM**

KPP PRATAMA SURABAYA GUBENG  
Jl. Sumatera No.22-24, Gubeng, Surabaya, Jawa Timur  
Tanggal Pelaksanaan : 13 Januari 2025 – 09 Mei 2025

**Oleh:**

Puput Ayu Setiawati                      3322600004

Surabaya, 13 Juni 2025

**Menyetujui:**

Pembimbing Kerja Praktek

**Fitrah Maharani Humaira, M.Kom**  
**NIP. 198906202022032008**

**Mengetahui:**

Pembimbing Lapangan  
KPP Pratama Surabaya Gubeng

**Diah Kusuma Sugiarto**  
**NIP. 810450527**

Halaman ini sengaja dikosongkan

## ABSTRAK

Kantor Pelayanan Pajak (KPP) Pratama Surabaya Gubeng merupakan instansi pemerintah yang memberikan layanan perpajakan bagi masyarakat. Ulasan masyarakat terhadap layanan tersebut dapat diamati melalui opini digital yang tersebar di berbagai *platform*, sehingga memudahkan untuk mengetahui persepsi dan tingkat kepuasan masyarakat. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap layanan KPP Pratama Surabaya Gubeng menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM). Data diperoleh melalui metode *web scraping* dari Google Maps dan kuesioner Google Forms. Seluruh data kemudian melalui tahap preprocessing, normalisasi, dan *balancing* sebelum dilakukan pemodelan. Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* memiliki performa lebih baik secara konsisten pada berbagai porsi data *training*, khususnya pada porsi 80% dan 90%, dengan nilai akurasi dan *F1-score* lebih tinggi dibandingkan SVM. Sedangkan SVM hanya unggul di porsi *training* 70%, namun performanya menurun saat porsi data meningkat. Temuan ini membuktikan bahwa algoritma berbasis probabilistik seperti *Naive Bayes* lebih sesuai untuk karakteristik data ulasan layanan perpajakan di KPP Pratama Surabaya Gubeng. Hasil klasifikasi ini diharapkan dapat menjadi bahan evaluasi dan pengambilan keputusan dalam peningkatan kualitas layanan berbasis opini masyarakat.

**Kata kunci:** Pajak, Sentimen, *Naive Bayes*, SVM, Google Maps, Google Forms.

Halaman ini sengaja dikosongkan

## KATA PENGANTAR

Dengan segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, atas limpahan rahmat, karunia, serta hidayah-Nya, sehingga penulis dapat menyelesaikan kegiatan Kerja Praktik (KP) di KPP Pratama Surabaya Gubeng dan penyusunan laporan Kerja Praktik dengan judul “Klasifikasi Sentimen Masyarakat Terhadap Layanan KPP Pratama Surabaya Gubeng Berbasis Naive Bayes dan SVM” dengan baik dan tepat waktu. Pelaksanaan Kerja Praktik ini menjadi bagian penting dalam menunjang proses pembelajaran dan persiapan menghadapi dunia kerja, karena memberikan kesempatan kepada mahasiswa untuk menerapkan teori yang telah dipelajari di bangku perkuliahan dalam menghadapi permasalahan nyata di lingkungan kerja sehingga memberikan banyak pengalaman dan wawasan yang sangat berharga bagi penulis. Adapun kegiatan ini dilaksanakan dalam rangka memenuhi salah satu persyaratan wajib bagi mahasiswa Politeknik Elektronika Negeri Surabaya pada semester 6 Program Studi Sarjana Terapan Sains Data Terapan.

Dalam kesempatan penyusunan ini, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah memberikan dukungan, bimbingan, serta saran yang membangun, sehingga laporan Kerja Praktik ini dapat diselesaikan tepat waktu. Ucapan terima kasih ini penulis tujukan kepada:

1. **Tuhan Yang Maha Esa**, karena limpahan rahmat, petunjuk, kemudahan, serta keyakinan yang diberikan, sehingga penulis dapat menyelesaikan kegiatan Kerja Praktik di KPP Pratama Surabaya Gubeng dengan baik dan lancar.
2. **Kedua orang tua beserta keluarga besar** atas doa, dukungan moral dan moril, serta bantuan material yang tak ternilai harganya, yang senantiasa memberikan semangat selama proses pelaksanaan Kerja Praktik hingga penyusunan laporan ini.
3. **Bapak Dr. Ing. Ir. Arif Irwansyah, S.T., M.Eng.** selaku Direktur Politeknik Elektronika Negeri Surabaya.
4. **Bapak Isbat Uzzin Nadhori S.Kom., M.T** selaku Ketua Program Studi Sarjana Terapan Sains Data Terapan.
5. **Ibu Alfi Fadliana S.Si., M.Stat** selaku Koordinator Kerja Praktik Program Studi Sarjana Terapan Sains Data Terapan.
6. **Bapak Ahmad Syauqi Ahsan, S.Kom., MT. Dan Ibu Fitrah Maharani Humaira, M.Kom** selaku Dosen Pembimbing Kerja Praktik yang telah memberikan arahan serta bimbingan selama periode Kerja Praktik.

7. **Bapak Sakdun** selaku Kepala Kantor yang telah memberikan kesempatan bagi penulis untuk melakukan Kerja Praktik di KPP Pratama Surabaya Gubeng.
8. **Ibu Diah Kusuma Sugiarto** selaku pembimbing lapangan Kerja Praktik di KPP Pratama Surabaya Gubeng yang telah memberikan waktu serta ilmu selama Kerja Praktik.
9. **Seluruh staff dan pegawai** di KPP Pratama Surabaya Gubeng yang telah memberikan waktu serta ilmu selama Kerja Praktik.
10. **Teman-teman kelas D4 Sains Data Terapan** yang saling mendukung, mengingatkan, memberi motivasi, dan mendoakan.
11. **Semua pihak lain** yang terlibat dalam penyusunan laporan Kerja Praktik ini.

Penulis menyadari bahwa laporan Kerja Praktik ini masih memiliki berbagai kekurangan, baik dalam hal materi maupun teknik penyajiannya. Hal ini tidak terlepas dari keterbatasan pengetahuan dan pengalaman yang dimiliki penulis. Oleh karena itu, segala bentuk kritik dan saran yang bersifat membangun sangat penulis harapkan demi penyempurnaan laporan di masa mendatang. Semoga laporan ini dapat memberikan manfaat, baik bagi pembaca secara umum maupun bagi penulis secara pribadi.

Surabaya, 13 Juni 2025

Penulis



## DAFTAR ISI

|   |       |
|---|-------|
| HALAMAN PENGESAHAN KERJA PRAKTIK.....                     | iii   |
| HALAMAN PENGESAHAN KERJA PRAKTIK.....                     | vii   |
| ABSTRAK.....  | xi    |
| KATA PENGANTAR.....                                       | xiii  |
| DAFTAR ISI .....  | xv    |
| DAFTAR GAMBAR .....                                       | xxi   |
| DAFTAR TABEL.....   | xxiii |
| BAB 1 PENDAHULUAN.....                                    | 1     |
| 1.1. Latar Belakang.....                                  | 1     |
| 1.2. Perumusan Masalah .....                              | 2     |
| 1.3. Tujuan dan Manfaat .....                             | 2     |
| 1.3.1. Tujuan .....                                       | 2     |
| 1.3.1.1. Tujuan Umum .....                                | 2     |
| 1.3.1.2. Tujuan Khusus.....                               | 3     |
| 1.3.2. Manfaat.....                                       | 3     |
| 1.3.2.1. Bagi Mahasiswa .....                             | 3     |
| 1.3.2.2. Bagi Politeknik Elektronika Negeri Surabaya .... | 4     |
| 1.3.2.3. Bagi KPP Pratama Surabaya Gubeng .....           | 4     |
| 1.4. Ruang Lingkup Pembahasan .....                       | 5     |
| 1.5. Sistematika Penulisan.....                           | 5     |
| BAB 2 LANDASAN TEORI.....                                 | 9     |
| 2.1. Gambaran Umum Perusahaan .....                       | 9     |
| 2.1.1. Profil Perusahaan .....                            | 9     |
| 2.1.2. Sejarah Singkat Perusahaan .....                   | 9     |

|   |    |
|---|----|
| 2.1.2.1. Nilai-Nilai Kementrian Keuangan .....    | 12 |
| 2.1.3. Struktur Organisasi.....                   | 14 |
| 2.1.3.1. KPP Pratama Kelompok I.....              | 16 |
| 2.1.3.2. KPP Pratama Kelompok II.....             | 19 |
| 2.1.3.3. Visi, Misi, dan Tujuan Perusahaan .....  | 21 |
| 2.1.3.3.1. Visi Perusahaan .....                  | 21 |
| 2.1.3.3.2. Misi Perusahaan .....                  | 22 |
| 2.1.3.4. Tugas Pokok dan Fungsi KPP Pratama.....  | 22 |
| 2.1.4. Hak dan Wewenang.....                      | 23 |
| 2.1.4.1. Hak .....                                | 23 |
| 2.1.4.2. Wewenang .....                           | 24 |
| 2.1.5. Lokasi Perusahaan.....                     | 25 |
| 2.2. Landasan Teori .....                         | 25 |
| 2.2.1. Data Teks.....                             | 25 |
| 2.2.2. Python .....                               | 27 |
| 2.2.3. <i>Text Mining</i> .....                   | 29 |
| 2.2.4. Analisis Sentimen.....                     | 30 |
| 2.2.5. Google Maps.....                           | 32 |
| 2.2.6. Google Forms .....                         | 34 |
| 2.2.7. <i>Web Scraping</i> .....                  | 36 |
| 2.2.8. <i>Preprocessing Data</i> .....            | 38 |
| 2.2.9. <i>Naive Bayes</i> .....                   | 40 |
| 2.2.10. <i>Support Vector Machine (SVM)</i> ..... | 42 |
| 2.2.11. Evaluasi Model Klasifikasi .....          | 44 |
| 2.2.12. Uji Normalisasi Data .....                | 47 |
| 2.2.13. <i>Data Balancing Methods</i> .....       | 48 |

|  |    |
|--|----|
| BAB 3 METODE PENELITIAN.....   | 51 |
| 3.1. Metode Pengumpulan Data .....   | 51 |
| 3.1.1. Metode Kepustakaan .....  | 51 |
| 3.1.2. Metode Web Scraping .....   | 51 |
| 3.1.3. Metode Wawancara.....   | 51 |
| 3.1.4. Metode Kuesioner .....  | 52 |
| 3.2. Prosedur Penelitian .....   | 53 |
| 3.3. Kontribusi.....   | 54 |
| 3.3.1. Membantu Tugas di Subbagian Umum dan Kepatuhan .....  | 55 |
| 3.3.2. Mempelajari Cara Pelayanan Pembuatan SPT .....  | 55 |
| 3.3.3. Blast WhatsApp Reminder SPT .....   | 55 |
| 3.3.4. Analisis Prediksi Risiko Wajib Pajak .....  | 56 |
| 3.3.5. Fuzzy Matching Data Restaurant Kecamatan Sukolilo .....   | 57 |
| 3.3.6. Penyusunan <i>Database</i> PPAT .....   | 58 |
| 3.3.7. Menginputkan Data Notaris untuk Pajak Pembuatan Akta Tanah (PPAT) .....                                       | 59 |
| 3.3.8. Input Laporan Hasil Kegiatan Pengumpulan Data ke Akses Pajak .....  | 59 |
| 3.3.9. Klasifikasi Sentimen Masyarakat Terhadap Layanan .....  | 60 |
| 3.4. Korelasi Kegiatan KP dengan Mata Kuliah.....  | 60 |
| BAB 4 ANALISA DAN PEMBAHASAN .....   | 63 |
| 4.1. Klasifikasi Sentimen Masyarakat Terhadap Layanan KPP Pratama Surabaya Gubeng Berbasis Naive Bayes dan SVM ..... | 63 |
| 4.2. Hasil dan Pembahasan.....   | 63 |
| 4.2.1. Dataset.....  | 63 |

|  |    |
|--|----|
| 4.2.1.1. Hasil Scraping Google Maps.....                               | 64 |
| 4.2.1.2. Hasil Kuesioner Google Forms .....                            | 64 |
| 4.2.2. Proses Preprocessing Data .....                                 | 65 |
| 4.2.2.1. Penggabungan Data .....                                       | 65 |
| 4.2.2.2. <i>Cleaning</i> .....   | 66 |
| 4.2.2.3. <i>Case Folding</i> .....                                     | 67 |
| 4.2.2.4. <i>Normalization</i> .....                                    | 67 |
| 4.2.2.5. <i>Tokenizing</i> .....                                       | 68 |
| 4.2.2.6. <i>Stopword Removal</i> .....                                 | 68 |
| 4.2.2.7. <i>Stemming</i> .....   | 69 |
| 4.2.3. Pemetaan Topik Ulasan .....                                     | 69 |
| 4.2.4. Ekplorasi Data.....   | 71 |
| 4.2.5. Pemodelan dan Ekperimen .....                                   | 73 |
| 4.2.5.1. Pelabelan Sentimen.....                                       | 73 |
| 4.2.5.2. Penyeimbangan Dataset .....                                   | 75 |
| 4.2.5.3. Klasifikasi Menggunakan Model Naive Bayes<br>(NB) .....       | 75 |
| 4.2.5.4. Klasifikasi Menggunakan Support Vector<br>Machine (SVM) ..... | 76 |
| 4.2.6. Perbandingan Kinerja Model Klasifikasi .....                    | 78 |
| 4.2.7. Visualisasi Frekuensi Kata dengan Word Cloud .....              | 80 |
| 4.2.7.1. Analisis Frekuensi Kata pada Topik ‘Pelayanan’<br>.....       | 82 |
| 4.2.7.2. Analisis Frekuensi Kata pada Topik ‘Tempat’ .                 | 83 |
| 4.2.5.1. Analisis Frekuensi Kata pada Topik ‘Overall’ .                | 85 |
| BAB V KESIMPULAN DAN SARAN .....                                       | 87 |
| 5.1. Kesimpulan.....   | 87 |

|   |    |
|---|----|
| 5.2. Saran .....  | 88 |
| 5.2.1. Kepada KPP Pratama Surabaya Gubeng .....             | 88 |
| 5.2.2. Kepada Politeknik Elektronika Negeri Surabaya .....  | 88 |
| 5.2.3. Kepada Penulis.....                                  | 89 |
| DAFTAR PUSTAKA.....   | 90 |
| LAMPIRAN .....  | 93 |
| Lampiran 1. Biodata Penulis 1 .....                         | 93 |
| Lampiran 2. Biodata Penulis 2 .....                         | 95 |
| Lampiran 3. Surat Pernyataan Penerimaan Kerja Praktik ..... | 81 |
| Lampiran 4. Sertifikat.....                                 | 82 |
| Lampiran 5. Form Bimbingan KP .....                         | 83 |
| Lampiran 6. Dokumentasi.....                                | 85 |
| Lampiran 7. File Program .....                              | 86 |

Halaman ini sengaja dikosongkan

## DAFTAR GAMBAR

|                      |   |    |
|----------------------|---|----|
| <b>Gambar 2. 1.</b>  | Logo KPP Pratama.....                               | 9  |
| <b>Gambar 2. 2.</b>  | Gedung KPP Pratama Surabaya Gubeng.....             | 10 |
| <b>Gambar 2. 3.</b>  | Ruang Pelayanan KPP Pratama Surabaya Gubeng .....   | 11 |
| <b>Gambar 2. 4.</b>  | Nilai-Nilai Kementerian Keuangan .....              | 13 |
| <b>Gambar 2. 5.</b>  | Struktur Organisasi DJP .....                       | 14 |
| <b>Gambar 2. 6.</b>  | Struktur Organisasi KPP.....                        | 14 |
| <b>Gambar 2. 7.</b>  | Lokasi KPP Pratama Surabaya Gubeng .....            | 25 |
| <b>Gambar 2. 8.</b>  | Visualisasi Data Teks.....                          | 26 |
| <b>Gambar 2. 9.</b>  | Logo Python .....                                   | 27 |
| <b>Gambar 2. 10.</b> | Ilustrasi Proses Text Mining .....                  | 29 |
| <b>Gambar 2. 11.</b> | Ilustrasi Hasil Analisis Sentimen .....             | 32 |
| <b>Gambar 2. 12.</b> | Logo Google Maps.....                               | 33 |
| <b>Gambar 2. 13.</b> | Logo Google Forms .....                             | 35 |
| <b>Gambar 2. 14.</b> | Ilustrasi Web Scraping untuk Machine Learning ..... | 36 |
| <b>Gambar 2. 15.</b> | Ilustrasi Teknik-Teknik Text Preprocessing .....    | 38 |
| <b>Gambar 2. 16.</b> | Diagram Konsep Algoritma Naive Bayes.....           | 40 |
| <b>Gambar 2. 17.</b> | Ilustrasi Konsep Support Vector Machine (SVM).....  | 42 |
| <b>Gambar 2. 18.</b> | Diagram Confusion Matrix.....                       | 45 |
| <b>Gambar 2. 19.</b> | Transformasi Uji Normalisasi data.....              | 47 |
| <b>Gambar 2. 20.</b> | Transformasi Uji Normalisasi data.....              | 49 |
| <br>                 |   |    |
| <b>Gambar 3. 1.</b>  | Flowchart Prosedur Penelitian .....                 | 53 |
| <b>Gambar 3. 2.</b>  | Prediksi Risiko Wajib Pajak .....                   | 56 |
| <b>Gambar 3. 3.</b>  | Program Fuzzy Matcing .....                         | 57 |
| <b>Gambar 3. 4.</b>  | Dahboard PPAT.....                                  | 58 |
| <br>                 |   |    |
| <b>Gambar 4. 1.</b>  | Chart Pemetaan Topik.....                           | 71 |
| <b>Gambar 4. 2.</b>  | Visualisasi Distribusi Rating Bintang.....          | 71 |
| <b>Gambar 4. 3.</b>  | Jumlah Data Ulasan per Tahun .....                  | 72 |
| <b>Gambar 4. 4.</b>  | Chart Distribusi Sentimen .....                     | 74 |
| <b>Gambar 4. 5.</b>  | Confussion Matrix NB .....                          | 75 |
| <b>Gambar 4. 6.</b>  | Evaluasi Performansi Model Naive Bayes.....         | 76 |
| <b>Gambar 4. 7.</b>  | Confussion Matrix SVM .....                         | 77 |

|                      |  |           |
|----------------------|--|-----------|
| <b>Gambar 4. 8.</b>  | <b>Evaluasi Performansi Model SVM.....</b> | <b>78</b> |
| <b>Gambar 4. 9.</b>  | <b>Wordcloud .....</b>                     | <b>80</b> |
| <b>Gambar 4. 10.</b> | <b>Frekuensi Kata .....</b>                | <b>81</b> |
| <b>Gambar 4. 11.</b> | <b>Wordcloud Pelayanan.....</b>            | <b>82</b> |
| <b>Gambar 4. 12.</b> | <b>Frekuensi Kata Topik Pelayanan.....</b> | <b>83</b> |
| <b>Gambar 4. 13.</b> | <b>WordCloud Tempat.....</b>               | <b>84</b> |
| <b>Gambar 4. 14.</b> | <b>Frekuensi Kata Topik Tempat.....</b>    | <b>84</b> |
| <b>Gambar 4. 15.</b> | <b>WordCloud Overall.....</b>              | <b>85</b> |
| <b>Gambar 4. 16.</b> | <b>Frekuensi Kata Topik Overall.....</b>   | <b>86</b> |



## DAFTAR TABEL

|                    |  |           |
|--------------------|--|-----------|
| <b>Tabel 4. 1.</b> | <b>Info Data .....</b>                     | <b>66</b> |
| <b>Tabel 4. 2.</b> | <b>Sampel Data Responden .....</b>         | <b>66</b> |
| <b>Tabel 4. 3.</b> | <b>Ulasan Duplikat.....</b>                | <b>67</b> |
| <b>Tabel 4. 4.</b> | <b>Hasil Tokenizing .....</b>              | <b>68</b> |
| <b>Tabel 4. 5.</b> | <b>Hasil Stemming.....</b>                 | <b>69</b> |
| <b>Tabel 4. 6.</b> | <b>Pemetaan Topik.....</b>                 | <b>70</b> |
| <b>Tabel 4. 7.</b> | <b>Perbandingan Metrik Kinerja .....</b>   | <b>78</b> |
| <b>Tabel 4. 8.</b> | <b>Akurasi Sentimen NB Per Topik .....</b> | <b>79</b> |
| <b>Tabel 4. 9.</b> | <b>Akurasi Sentimen SVM Per Topik.....</b> | <b>79</b> |

Halaman ini sengaja dikosongkan

# **BAB 1**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Mahasiswa yang memiliki kemampuan di bidang teoretis maupun praktis menjadi kebutuhan penting bagi berbagai instansi, termasuk dalam sektor pemerintahan. Untuk mendukung tercapainya tujuan tersebut, Politeknik Elektronika Negeri Surabaya (PENS) menyelenggarakan program Kerja Praktik yang dilaksanakan pada semester kelima bagi mahasiswa jenjang D4. Program ini bertujuan untuk memberikan pengalaman langsung kepada mahasiswa mengenai dunia kerja, sekaligus memperkuat hubungan antara perguruan tinggi dengan instansi mitra. Kerja Praktik juga menjadi sarana bagi mahasiswa untuk mengimplementasikan ilmu yang telah diperoleh di bangku kuliah, khususnya dalam bidang Pemrosesan Data, Eksplorasi dan Visualisasi Data, serta Analisis Media Sosial, yang relevan dengan kebutuhan dunia kerja saat ini.

Kantor Pelayanan Pajak (KPP) Pratama Surabaya Gubeng merupakan salah satu unit pelaksana teknis di bawah Direktorat Jenderal Pajak yang memiliki peran strategis dalam menghimpun penerimaan negara di wilayah Surabaya. Terletak di pusat kota yang dinamis, KPP Pratama Surabaya Gubeng melayani berbagai jenis wajib pajak, mulai dari pelaku usaha, profesional, hingga perusahaan skala besar. Dengan volume pelayanan yang tinggi, KPP ini terus berupaya meningkatkan efisiensi dan efektivitas kinerjanya melalui optimalisasi penggunaan teknologi informasi, modernisasi administrasi perpajakan, serta pengelolaan basis data yang akurat dan terintegrasi untuk menunjang pelayanan publik yang prima.

Melalui pelaksanaan Kerja Praktik di KPP Pratama Surabaya Gubeng, mahasiswa memperoleh kesempatan untuk mengembangkan kompetensi teknis dan praktis, khususnya dalam bidang administrasi data dan pelayanan perpajakan. Keterlibatan mahasiswa dalam kegiatan operasional diharapkan dapat meningkatkan pemahaman mereka terhadap tata kelola administrasi pemerintahan, penggunaan sistem informasi perpajakan, serta pentingnya pelayanan berbasis data

dalam mendukung kinerja instansi. Selain itu, pengalaman ini juga membekali mahasiswa dengan keterampilan yang diperlukan untuk menghadapi tantangan dunia kerja yang semakin kompetitif di masa depan.

## **1.2. Perumusan Masalah**

Dalam pelaksanaan Kerja Praktik ini dan berdasarkan latar belakang yang telah diuraikan, permasalahan yang ingin diselesaikan adalah dengan bagaimana cara mengumpulkan serta memproses data sentimen masyarakat terkait layanan KPP Pratama Surabaya Gubeng yang diperoleh melalui *platform* media sosial dan kuesioner. Selain itu, dilakukan pembangunan model klasifikasi sentimen menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) untuk mengelompokkan sentimen menjadi kategori positif dan negatif. Selanjutnya, akan dibandingkan performa kedua model klasifikasi tersebut guna mengetahui algoritma mana yang memberikan hasil terbaik dalam analisis sentimen.

## **1.3. Tujuan dan Manfaat**

### **1.3.1. Tujuan**

Politeknik Elektronika Negeri Surabaya (PENS) senantiasa berupaya membekali mahasiswa dengan pengalaman langsung didunia kerja agar mampu memahami permasalahan riil yang terjadi. Melalui pelaksanaan Kerja Praktik di KPP Pratama Surabaya Gubeng, diharapkan dapat mencapai tujuan sebagai berikut:

#### **1.3.1.1. Tujuan Umum**

1. Memberikan pengalaman langsung kepada mahasiswa mengenai proses administrasi, pelayanan, dan pengelolaan data di lingkungan pemerintahan, khususnya di KPP Pratama Surabaya Gubeng.
2. Mengembangkan kemampuan mahasiswa dalam menerapkan ilmu yang diperoleh di perkuliahan ke dalam situasi nyata di lingkungan kerja.

3. Menyediakan gambaran nyata bagi mahasiswa mengenai penerapan teknologi informasi dan *data science* dalam upaya peningkatan kualitas layanan di sektor perpajakan.
4. Membantu KPP Pratama Surabaya Gubeng dalam mengoptimalkan pelayanan publik berbasis data melalui analisis sentimen masyarakat terhadap layanan yang diberikan.

#### **1.3.1.2. Tujuan Khusus**

1. Mengumpulkan data sentimen masyarakat terkait layanan KPP Pratama Surabaya Gubeng melalui *platform* media sosial dan kuesioner.
2. Melakukan proses *preprocessing data*, pengolahan, dan eksplorasi data sentimen untuk keperluan analisis.
3. Membangun model klasifikasi sentimen menggunakan algoritma *Naive Bayes*.
4. Membangun model klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM).
5. Melakukan evaluasi dan perbandingan kinerja antara model *Naive Bayes* dan SVM dalam mengklasifikasikan sentimen masyarakat.
6. Memberikan rekomendasi hasil analisis sentimen yang dapat digunakan oleh KPP Pratama Surabaya Gubeng dalam upaya meningkatkan kualitas pelayanan publik.

#### **1.3.2. Manfaat**

Dalam Pelaksanaan Kerja Praktik ini, dapat memberikan manfaat yang didapatkan antara lain:

##### **1.3.2.1. Bagi Mahasiswa**

1. Mahasiswa dapat memahami secara lebih mendalam realita ilmu yang telah dipelajari di perkuliahan ketika dihadapkan dengan kondisi dan tantangan langsung di lapangan.
2. Menyiapkan diri untuk menyesuaikan dengan lingkungan industri pada masa yang akan datang.

3. Mahasiswa mampu mengintegrasikan pengetahuan akademik yang diperoleh selama perkuliahan dengan pengalaman kerja nyata di dunia industri.
4. Menambah pengalaman nyata di lingkungan kerja pemerintahan khususnya dalam bidang administrasi data dan pelayanan perpajakan.
5. Menambah wawasan tentang implementasi teknologi informasi dan data science dalam mendukung layanan publik.
6. Meningkatkan keterampilan teknis mahasiswa dalam pengolahan data, analisis media sosial, serta penerapan algoritma klasifikasi seperti *Naive Bayes* dan SVM.
7. Mengasah kemampuan mahasiswa dalam menyusun laporan Kerja Praktik yang sistematis dan berbasis analisis data aktual.

#### **1.3.2.2. Bagi Politeknik Elektronika Negeri Surabaya**

1. Mempererat hubungan kerja sama antara Politeknik Elektronika Negeri Surabaya dengan instansi pemerintah, khususnya KPP Pratama Surabaya Gubeng.
2. Menjadi media untuk mengimplementasikan kurikulum berbasis praktik yang relevan dengan kebutuhan industri dan pemerintahan.
3. Memberikan kontribusi dalam bentuk solusi berbasis data science yang dapat diaplikasikan di dunia kerja nyata.
4. Menjadi bahan evaluasi dan pengembangan kurikulum agar sesuai dengan perkembangan teknologi serta kebutuhan dunia kerja saat ini.

#### **1.3.2.3. Bagi KPP Pratama Surabaya Gubeng**

1. Sebagai sarana untuk mengetahui kualitas pendidikan di Perguruan Tinggi Negeri, khususnya Politeknik Elektronika Negeri Surabaya.
2. Sebagai sarana untuk mengenalkan teknologi industri khususnya informasi pada dunia pendidikan.
3. Mendapatkan hasil analisis sentimen masyarakat terkait layanan perpajakan yang dapat digunakan sebagai bahan evaluasi dan peningkatan kualitas pelayanan.

4. Meningkatkan efektivitas pengelolaan opini masyarakat di media sosial melalui pemanfaatan teknologi klasifikasi sentimen.
5. Dengan program ini akan bermanfaat bagi perusahaan untuk menilai kemampuan mahasiswa sebagai acuan dalam penerimaan lulusan perguruan tinggi.

#### **1.4. Ruang Lingkup Pembahasan**

Ruang lingkup pembahasan dalam laporan kerja praktik ini difokuskan pada pengumpulan dan pengolahan data sentimen masyarakat yang diperoleh dari *platform* media sosial dan kuesioner terkait layanan di KPP Pratama Surabaya Gubeng. Data yang dikumpulkan terbatas pada komentar atau opini masyarakat yang relevan dengan layanan perpajakan di lingkungan KPP Pratama Surabaya Gubeng selama periode tertentu.

Selanjutnya, dilakukan penerapan algoritma klasifikasi sentimen menggunakan metode *Naive Bayes* dan *Support Vector Machine* (SVM) untuk mengkategorikan sentimen menjadi dua kelas, yaitu positif dan negatif. Proses ini mencakup tahapan *preprocessing data*, pelabelan sentimen, pelatihan model, pengujian, serta evaluasi kinerja masing-masing model. Pembahasan juga mencakup perbandingan hasil klasifikasi dari kedua algoritma tersebut untuk menentukan metode mana yang paling efektif dalam mengklasifikasikan sentimen masyarakat. Ruang lingkup kerja praktik ini tidak mencakup perancangan sistem informasi, pembuatan visualisasi, maupun analisis data perpajakan secara menyeluruh, melainkan hanya terbatas pada analisis sentimen berbasis teks dari media sosial. Hasil dari analisis ini diharapkan dapat memberikan gambaran umum tentang opini masyarakat terhadap layanan KPP Pratama Surabaya Gubeng serta menjadi rekomendasi awal bagi pihak instansi dalam upaya meningkatkan kualitas layanan publik berbasis data.

#### **1.5. Sistematika Penulisan**

Sebagai upaya memberikan gambaran yang jelas dan terstruktur mengenai isi laporan ini, penulis menyusun sistematika penulisan yang dibagi menjadi beberapa bab. Setiap bab berisi pembahasan yang

saling terkait dan berfokus pada tujuan serta hasil yang ingin dicapai dalam kerja praktik ini. Adapun sistematika penulisan laporan ini adalah sebagai berikut:

### **1. BAB I: PENDAHULUAN**

Bab ini menyajikan latar belakang, perumusan masalah tentang kondisi yang perlu di evaluasi di tempat Kerja Praktik, tujuan dan manfaat, ruang lingkup pembahasan, serta sistematika pembahasan kerja praktik. Di bagian ini, penulis menjelaskan konteks yang mendasari penelitian dan menjelaskan arah serta harapan dari proyek yang dilakukan.

### **2. BAB II: LANDASAN TEORI**

Bab ini menjelaskan penjelasan mengenai profil singkat perusahaan tempat pelaksanaan Kerja Praktik, yang mencakup visi, misi, serta tujuan dari KPP Pratama Surabaya Gubeng. Tidak hanya itu, nilai-nilai kebiasaan positif yang diterapkan oleh para karyawan, seperti sikap disiplin, kerjasama, dan etika kerja, turut dijelaskan sebagai bagian dari budaya kerja di lingkungan perusahaan. Selain itu, dijelaskan pula struktur organisasi perusahaan yang dilengkapi dengan bagan atau diagram untuk memperjelas susunan unit kerja yang ada. Uraian mengenai hak, dan wewenang dari masing-masing unit kerja dalam struktur organisasi juga dijabarkan secara rinci. Bab ini turut memuat informasi tentang lokasi perusahaan yang dilengkapi dengan peta. Selain itu, bab ini juga memuat teori-teori dasar yang relevan dengan topik kerja praktik, seperti konsep klasifikasi sentimen, algoritma *Naive Bayes*, algoritma *Support Vector Machine* (SVM), serta metode *web scraping*. Selain itu, bab ini juga memuat hasil-hasil penelitian terdahulu yang berkaitan sebagai acuan.

### **3. BAB III: METODE PENELITIAN**

Bab ini menjelaskan metode yang digunakan dalam pelaksanaan kegiatan Kerja Praktik (KP) dan penelitian yang dilakukan selama KP berlangsung. Penjabaran dalam bab ini bertujuan untuk memberikan gambaran yang sistematis mengenai proses pengumpulan data, pelaksanaan penelitian, kontribusi selama KP,



serta hubungan antara kegiatan KP dengan mata kuliah yang telah dipelajari di bangku perkuliahan.

#### **4. BAB IV: ANALISIS DAN PEMBAHASAN**

Bab ini menguraikan secara rinci aktivitas yang telah dilakukan selama pelaksanaan Kerja Praktik. Bab ini diawali dengan penjelasan umum mengenai aktivitas kerja praktik, dilanjutkan dengan analisis sistem, dari proses pengambilan data, implementasi algoritma klasifikasi sentimen, serta pembahasan hasil perbandingan performa model *Naive Bayes* dan *Support Vector Machine* (SVM). Bab ini juga membahas kelebihan dan kekurangan sistem yang diterapkan, serta memberikan usulan perbaikan yang dapat diimplementasikan untuk meningkatkan kualitas pelayanan di KPP Pratama Surabaya Gubeng.

#### **5. BAB V: KESIMPULAN DAN SARAN**

Bab ini menyajikan kesimpulan dari hasil analisis dan pembahasan yang telah dilakukan pada bab sebelumnya, disajikan secara singkat, jelas, dan langsung mengacu pada ketercapaian tujuan dan manfaat yang telah dirumuskan di Bab I. Selain itu, disampaikan pula saran-saran yang bersifat membangun, baik untuk pihak KPP Pratama Surabaya Gubeng maupun bagi mahasiswa.

Halaman ini sengaja dikosongkan

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1. Gambaran Umum Perusahaan**

##### **2.1.1. Profil Perusahaan**



**Gambar 2. 1.** Logo KPP Pratama

**Sumber:** Kompas TV. (2023, 24 Februari). *Logo Direktorat Jenderal Pajak.*

Nama Perusahaan : KPP Pratama Surabaya Gubeng  
Alamat : Jalan Sumatera No. 22-24, Surabaya  
Telp/Fax : +62 31- 5036962 / +62 31-5031566  
Website/Email : [www.pajak.go.id](http://www.pajak.go.id) / kpp.606@pajak.go.id  
Instagram : @pajaksbygubeng  
YouTube : @kpppratamasurabayagubeng9131

##### **2.1.2. Sejarah Singkat Perusahaan**

Direktorat Jenderal Pajak (DJP) merupakan salah satu unsur pelaksana di lingkungan Kementerian Keuangan Republik Indonesia yang memiliki peran strategis dalam upaya penghimpunan penerimaan negara dari sektor perpajakan. Keberadaan DJP menjadi pilar utama dalam mendukung pembiayaan pembangunan nasional, penyelenggaraan pemerintahan, dan pelayanan kepada masyarakat. Dalam menjalankan tugas pokok dan fungsinya, DJP bertanggung jawab menyusun serta melaksanakan kebijakan dan standarisasi teknis di bidang perpajakan. Untuk memastikan kelancaran pelaksanaan

tugas tersebut di berbagai wilayah, DJP membentuk unit-unit pelaksana teknis yang tersebar di seluruh Indonesia.

Seiring berjalannya waktu, sejak tahun 2002, DJP secara bertahap melakukan modernisasi sistem dan struktur organisasi menuju instansi yang berorientasi pada fungsi. Modernisasi ini diawali dengan penggabungan Kantor Pelayanan Pajak Konvensional dan Kantor Pemeriksaan dan Penyidikan Pajak. Pada tahun yang sama, dibentuk pula dua KPP Wajib Pajak Besar atau *Large Tax Office* (LTO). Satu tahun kemudian, tepatnya tahun 2003, dibentuk sepuluh KPP Khusus, dilanjutkan pada tahun 2004 dengan pendirian KPP Madya atau *Medium Tax Office* (MTO). Selanjutnya, dua tahun kemudian, yaitu mulai tahun 2006 hingga 2008, dibuka KPP Modern yang lebih dikenal sebagai KPP Pratama atau *Small Tax Office* (STO), yang menjadi KPP terbanyak dengan cakupan layanan Wajib Pajak terbesar di seluruh wilayah Indonesia.



**Gambar 2. 2.** Gedung KPP Pratama Surabaya Gubeng

**Sumber:** Publikasi KPP Pratama Surabaya Gubeng

Salah satu KPP Pratama di bawah naungan Kantor Wilayah DJP Jawa Timur I adalah Kantor Pelayanan Pajak Pratama Surabaya Gubeng. Kantor ini berdiri sebagai wujud nyata komitmen pemerintah untuk terus meningkatkan pelayanan publik di sektor perpajakan, khususnya di wilayah Kota Surabaya dan sekitarnya. Secara resmi, KPP Pratama Surabaya Gubeng didirikan pada tahun 2003 seiring dengan program modernisasi administrasi perpajakan yang mulai digalakkan oleh DJP pada awal dekade 2000-an. Modernisasi tersebut bertujuan untuk

membangun sistem administrasi perpajakan yang lebih efektif, efisien, transparan, serta berorientasi pada pelayanan prima bagi Wajib Pajak.

Sebelum adanya modernisasi, pelayanan perpajakan di wilayah Surabaya ditangani secara umum oleh beberapa KPP tanpa pembagian wilayah kerja yang spesifik. Namun, seiring meningkatnya jumlah penduduk, aktivitas perekonomian, serta jumlah Wajib Pajak, pemerintah memandang perlu melakukan pemekaran wilayah kerja pelayanan perpajakan di Kota Surabaya. Salah satu hasil dari kebijakan tersebut adalah pembentukan KPP Pratama Surabaya Gubeng yang bertugas memberikan pelayanan administrasi perpajakan di kawasan Kecamatan Gubeng dan sekitarnya.



**Gambar 2. 3.** Ruang Pelayanan KPP Pratama Surabaya Gubeng  
**Sumber:** Publikasi KPP Pratama Surabaya Gubeng

KPP Pratama Surabaya Gubeng tidak hanya menyelenggarakan pelayanan administrasi perpajakan, tetapi juga melaksanakan pengawasan kepatuhan Wajib Pajak, penyuluhan perpajakan, serta penegakan hukum di bidang perpajakan sesuai ketentuan perundang-undangan yang berlaku. Sejak berdiri hingga saat ini, KPP Pratama Surabaya Gubeng terus mengalami berbagai perkembangan, baik dari sisi fasilitas, sistem pelayanan, maupun sumber daya manusia.

Selama lebih dari dua dekade berdiri, KPP Pratama Surabaya Gubeng terus melakukan inovasi layanan guna meningkatkan kepuasan Wajib Pajak. Salah satunya melalui pemanfaatan teknologi informasi dan platform digital, termasuk media sosial dan layanan daring, sebagai sarana komunikasi dua arah. Kantor ini juga aktif melaksanakan edukasi dan penyuluhan perpajakan kepada masyarakat, pelaku usaha, serta instansi pemerintahan di wilayah kerjanya.

#### **2.1.2.1. Nilai-Nilai Kementrian Keuangan**

Sebagai instansi strategis dalam menghimpun penerimaan negara, KPP Pratama Surabaya Gubeng memiliki beberapa budaya kerja sebagai nilai dasar dalam setiap aktivitas pelayanan, seperti:

1. Integritas  
Berpikir, berkata, berperilaku dan bertindak dengan baik dan benar serta memegang teguh kode etik dan prinsip-prinsip moral.  
Prilaku Utama: Bersikap jujur, disiplin, tulus, konsisten, dapat dipercaya, menjaga martabat, tidak melakukan hal-hal tercela dan militan dalam bekerja.
2. Profesionalisme  
Bekerja tuntas dan akurat atas dasar kompetensi terbaik dengan penuh tanggung jawab dan komitmen yang tinggi.  
Prilaku Utama: Mempunyai keahlian dan pengetahuan yang luas dan juga bekerja dengan hati
3. Sinergi  
Membangun dan memastikan hubungan kerjasama internal yang produktif serta kemitraan yang harmonis dengan para pemangku kepentingan, untuk menghasilkan karya yang bermanfaat dan berkualitas.  
Prilaku Utama: Memiliki prasangka baik, saling percaya dan menghormati, serta menemukan dan melaksanakan solusi terbaik
4. Pelayanan

Memberikan layanan yang memenuhi kepuasan pemangku kepentingan yang dilakukan dengan sepenuh hati, transparan, cepat, akurat dan aman.

Prilaku Utama: Melayani dengan berorientasi pada kepuasan pemangku kepentingan dan bersikap proaktif dan cepat tanggap

5. Kesempurnaan

Senantiasa melakukan upaya perbaikan di segala bidang untuk menjadi dan memberikan yang terbaik.

Prilaku Utama: Melakukan perbaikan terus menerus dan juga mengembangkan inovasi dan kreativitas

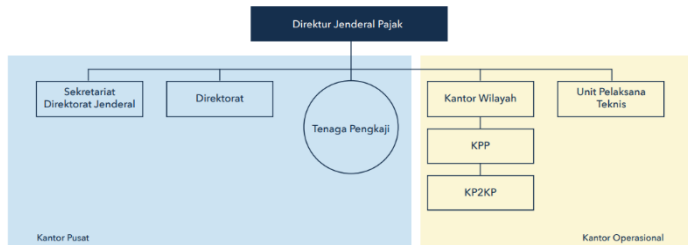


**Gambar 2. 4.** Nilai-Nilai Kementerian Keuangan

**Sumber:** Kementerian Keuangan RI. (2024). *Nilai-Nilai Kemenkeu RI*.

Budaya kerja tersebut didukung dengan penerapan standar pelayanan minimal, sistem pengawasan internal yang ketat, serta peningkatan kapasitas sumber daya manusia melalui berbagai program pelatihan dan pengembangan kompetensi. Dengan sejarah panjangnya sejak tahun 2003 hingga kini, KPP Pratama Surabaya Gubeng terus berkomitmen menjadi unit pelaksana teknis perpajakan yang modern, profesional, dan responsif terhadap perkembangan kebutuhan masyarakat serta teknologi informasi, guna mendukung tercapainya target penerimaan negara dan peningkatan kepatuhan perpajakan.

### 2.1.3. Struktur Organisasi



**Gambar 2. 5.** Struktur Organisasi DJP  
**Sumber:** KPP Pratama Surabaya Gubeng

Struktur Organisasi Kantor Pelayanan Pajak Pratama sesuai PMK 206.02/PMK.01/2014 yang mulai berlaku per 11 Juli 2022.



**Gambar 2. 6.** Struktur Organisasi KPP  
**Sumber:** KPP Pratama Surabaya Gubeng

Berdasarkan Peraturan Menteri Keuangan Republik Indonesia nomor 184/PMK.01/2020 tentang perubahan atas peraturan menteri keuangan nomor 210/pmk.01/2017 tentang organisasi dan tata kerja instansi vertikal direktorat jenderal pajak. Dalam melaksanakan tugas KPP Pratama menyelenggarakan fungsi:



1. Analisis, penjabaran, dan pencapaian target penerimaan pajak.
2. Penguasaan data dan informasi subjek dan objek pajak dalam wilayah wewenang KPP.
3. Pelayanan, edukasi, pendaftaran, dan pengelolaan pelaporan wajib pajak.
4. Pendaftaran wajib Pajak, objek pajak, dan penghapusan nomor pokok wajib pajak.
5. Pengukuhan dan pencabutan pengukuhan pengusaha kena pajak.
6. Pemberian dan/ atau penghapusan nomor objek pajak secara jabatan.
7. Penyelesaian tindak lanjut pengajuan/pencabutan permohonan wajib pajak maupun masyarakat.
8. Pengawasan, pemeriksaan, penilaian, dan penagihan pajak.
9. Pendataan, pemetaan wajib pajak dan objek pajak, dan pengenaan.
10. Penetapan, penerbitan, dan/ atau pembetulan produk hukum dan produk layanan perpajakan.
11. Pengawasan dan pemantauan tindak lanjut pengampunan pajak.
12. Penjaminan kualitas data hasil perekaman dan hasil identifikasi data internal dan eksternal.
13. Pemutakhiran basis data perpajakan.
14. Pengurangan pajak bumi dan bangunan
15. Pengelolaan kinerja dan pengelolaan risiko.
16. Pelaksanaan dan pemantauan kepatuhan internal.
17. Penatausahaan dan pengelolaan piutang pajak.
18. Pelaksanaan tindak lanjut kerja sama perpajakan.
19. Pengelolaan dokumen perpajakan dan nonperpajakan.
20. Pelaksanaan administrasi kantor.

Dalam pelaksanaan tugas dan fungsinya, Kantor Pelayanan Pajak (KPP) Pratama di lingkungan Direktorat Jenderal Pajak dikelompokkan ke dalam beberapa kategori untuk memastikan efektivitas dan efisiensi pelayanan kepada masyarakat. Secara

umum, KPP Pratama dibagi menjadi dua kelompok utama berdasarkan karakteristik wajib pajak dan cakupan wilayah kerja yang dilayaninya. Pengelompokan ini bertujuan agar pengelolaan administrasi perpajakan dapat berjalan lebih terarah, sesuai dengan kompleksitas permasalahan perpajakan yang dihadapi masing-masing wilayah serta karakter wajib pajaknya. Selain itu, pembagian kelompok ini juga menjadi dasar dalam penyusunan strategi pelayanan, sistem pengawasan, hingga pelaksanaan penyuluhan perpajakan, sehingga seluruh proses dapat berlangsung optimal dan sejalan dengan target penerimaan negara yang telah ditetapkan.

#### **2.1.3.1. KPP Pratama Kelompok I**

Untuk mendukung kelancaran pelaksanaan tugas dan pelayanan kepada masyarakat, Kantor Pelayanan Pajak (KPP) Pratama Kelompok I memiliki susunan organisasi yang terdiri dari beberapa subbagian dan seksi. Setiap bagian memiliki peran, fungsi, dan tanggung jawab yang saling berkaitan dalam rangka mengoptimalkan pelayanan perpajakan. Adapun struktur organisasi di KPP Pratama Kelompok I terdiri dari:

1. Subbagian Umum dan Kepatuhan Internal.
2. Seksi Penjaminan Kualitas Data.
3. Seksi Pelayanan.
4. Seksi Pemeriksaan, Penilaian, dan Penagihan.
5. Seksi Pengawasan I.
6. Seksi Pengawasan II.
7. Seksi Pengawasan III.
8. Seksi Pengawasan IV.
9. Seksi Pengawasan V.
10. Seksi Pengawasan VI.
11. Kelompok Jabatan Fungsional.

Setiap subbagian dan seksi di lingkungan KPP Pratama Kelompok I memiliki tugas pokok dan fungsi yang berbeda-beda, namun saling mendukung untuk mencapai tujuan

organisasi. Berikut adalah uraian mengenai tugas dan tanggung jawab dari masing-masing subbagian dan seksi yang ada di dalam struktur organisasi KPP Pratama Kelompok I:

1. Subbagian Umum dan Kepatuhan Internal mempunyai tugas dalam mengelola urusan kepegawaian, administrasi keuangan, tata usaha, rumah tangga kantor, serta pengelolaan kinerja. Selain itu, subbagian ini juga bertanggung jawab dalam pemantauan pengendalian intern, pengujian kepatuhan, manajemen risiko, internalisasi kepatuhan, penyusunan laporan, pengelolaan dokumen non-perpajakan, serta memberikan dukungan teknis bagi pelaksanaan tugas-tugas kantor.
2. Seksi Penjaminan Kualitas Data mempunyai tugas melakukan analisis, penjabaran, serta pengelolaan kegiatan penjaminan kualitas data melalui perencanaan, pengumpulan, pengolahan, dan penyajian data serta informasi perpajakan. Seksi ini juga mengelola perekaman dokumen perpajakan, kerja sama perpajakan, serta menjamin kualitas data terkait intensifikasi dan ekstensifikasi perpajakan, distribusi dan tindak lanjut data, penatausahaan dokumen pembangunan data, serta pengelolaan administrasi produk hukum dan pengolahan data perpajakan, termasuk penyusunan monografi fiskal dan melakukan pengelolaan administrasi produk hukum dan produk pengolahan data perpajakan.
3. Seksi Pelayanan mempunyai tugas melakukan analisis, penjabaran, dan pengelolaan layanan perpajakan yang berkualitas serta memastikan wajib pajak memahami hak dan kewajibannya. Tugas tersebut meliputi edukasi dan konsultasi perpajakan, pengelolaan registrasi perpajakan, penerimaan dan pengolahan SPT, penanganan permohonan, saran, pengaduan, serta surat lainnya dari wajib pajak maupun masyarakat. Selain itu, seksi ini juga bertugas melakukan pemenuhan hak wajib pajak, penatausahaan dokumen, serta administrasi penetapan dan penerbitan produk hukum serta layanan perpajakan.

4. Seksi Pemeriksaan, Penilaian, dan Penagihan mempunyai tugas melakukan analisis, penjabaran, dan pengelolaan kegiatan dalam rangka pencapaian target penerimaan pajak. Kegiatan tersebut mencakup pelaksanaan pemeriksaan, penilaian properti, bisnis, dan aset tidak berwujud, tindakan penagihan, serta pengelolaan administrasi piutang pajak. Selain itu, seksi ini juga mengatur administrasi penetapan dan penerbitan produk hukum yang terkait dengan pemeriksaan, penilaian, dan penagihan pajak.
5. Seksi Pengawasan I, Seksi Pengawasan II, Seksi Pengawasan III, Seksi Pengawasan IV, Seksi Pengawasan V dan Seksi Pengawasan VI masing-masing mempunyai tugas melakukan analisis, penjabaran, dan pengelolaan dalam rangka memastikan wajib pajak mematuhi peraturan perundang-undangan perpajakan melalui perencanaan, pelaksanaan, dan tindak lanjut intensifikasi dan ekstensifikasi berbasis pendataan dan pemetaan (mapping) subjek dan objek pajak, penguasaan wilayah, pengamatan potensi pajak dan penguasaan informasi, pencarian, pengumpulan, pengolahan, penelitian, analisis, pemutakhiran, dan tindak lanjut data perpajakan, pengawasan dan pengendalian mutu kepatuhan kewajiban perpajakan wajib pajak, imbauan dan konseling kepada wajib pajak, pengawasan dan pemantauan tindak lanjut pengampunan pajak, serta melakukan pengelolaan administrasi penetapan dan penerbitan produk hukum dan produk pengawasan perpajakan. Ketentuan lebih lanjut mengenai pembagian dan penetapan tugas Seksi Pengawasan sebagaimana dimaksud pada ayat (5) ditetapkan dalam Keputusan Direktur Jenderal Pajak (DJP).
6. Kelompok Jabatan Fungsional merupakan pegawai yang menduduki jabatan fungsional tertentu dan bertugas memberikan dukungan teknis serta profesional sesuai dengan bidang keahliannya, seperti perpajakan,

pemeriksaan, penilaian, dan penagihan, berdasarkan ketentuan peraturan perundang-undangan yang berlaku.

#### **2.1.3.2. KPP Pratama Kelompok II**

Dalam rangka menjalankan tugas pelayanan, pengawasan, dan pemungutan pajak secara optimal, Kantor Pelayanan Pajak (KPP) Pratama Kelompok II memiliki susunan organisasi yang terdiri atas beberapa subbagian dan seksi. Setiap bagian dalam struktur ini memiliki peran strategis dalam mendukung pencapaian target penerimaan pajak, meningkatkan kualitas pelayanan kepada wajib pajak, serta memastikan kepatuhan perpajakan di wilayah kerjanya. Adapun struktur organisasi di KPP Pratama Kelompok II terdiri dari:

1. Subbagian Umum dan Kepatuhan Internal.
2. Seksi Penjaminan Kualitas Data.
3. Seksi Pelayanan.
4. Seksi Pemeriksaan, Penilaian, dan Penagihan.
5. Seksi Pengawasan I.
6. Seksi Pengawasan II.
7. Seksi Pengawasan III.
8. Seksi Pengawasan IV.
9. Seksi Pengawasan V.
10. Kelompok Jabatan Fungsional.

Setiap subbagian dan seksi di lingkungan KPP Pratama Kelompok II juga memiliki tugas pokok dan fungsi masing-masing, yang saling terintegrasi dalam mendukung kelancaran pelaksanaan tugas organisasi. Berikut adalah uraian mengenai tugas dan tanggung jawab dari masing-masing bagian tersebut:

1. Subbagian Umum dan Kepatuhan Internal mempunyai tugas melakukan urusan kepegawaian, keuangan, tata usaha, rumah tangga, pengelolaan kinerja, melakukan pemantauan pengendalian intern, pengujian kepatuhan dan manajemen risiko, internalisasi kepatuhan,

- penyusunan laporan, pengelolaan dokumen non perpajakan, serta dukungan teknis pelaksanaan tugas kantor.
2. Seksi Penjaminan Kualitas Data mempunyai tugas melakukan analisis dan pengelolaan dalam rangka penjaminan kualitas data perpajakan, yang mencakup perencanaan, pengumpulan, pengolahan, serta penyajian data dan informasi perpajakan. Seksi ini juga bertugas melakukan perekaman dokumen, tindak lanjut kerja sama, monitoring distribusi data, penyusunan monografi fiskal, hingga pengelolaan administrasi produk hukum dan produk pengolahan data.
  3. Seksi Pelayanan mempunyai tugas memberikan layanan perpajakan yang optimal kepada wajib pajak, termasuk edukasi, konsultasi, pengelolaan registrasi, penerimaan dan pengolahan surat pemberitahuan, serta pengelolaan permohonan, saran, pengaduan, dan surat lainnya. Seksi ini juga bertanggung jawab terhadap penatausahaan dokumen dan administrasi penerbitan produk layanan perpajakan.
  4. Seksi Pemeriksaan, Penilaian, dan Penagihan mempunyai tugas melakukan pemeriksaan pajak, penilaian properti, bisnis, dan aset tak berwujud, serta tindakan penagihan dan pengelolaan piutang pajak. Seksi ini juga bertugas dalam penatausahaan administrasi serta penerbitan produk hukum terkait pemeriksaan, penilaian, dan penagihan.
  5. Seksi Pengawasan I, Seksi Pengawasan II, Seksi Pengawasan III, Seksi Pengawasan IV, dan Seksi Pengawasan V Masing-masing seksi ini bertanggung jawab melakukan pengawasan terhadap wajib pajak, termasuk perencanaan, pelaksanaan, dan tindak lanjut intensifikasi dan ekstensifikasi pajak berbasis pendataan dan pemetaan subjek dan objek pajak. Selain itu, tugasnya mencakup pengamatan potensi, analisis, pemutakhiran data perpajakan, pengendalian mutu kepatuhan,

konseling wajib pajak, serta pengawasan tindak lanjut atas program pengampunan pajak. Ketentuan teknis mengenai pembagian tugas antar seksi pengawasan ditetapkan melalui Keputusan Direktur Jenderal Pajak.

6. Kelompok Jabatan Fungsional merupakan pegawai dengan jabatan fungsional tertentu yang bertugas memberikan dukungan teknis dan profesional sesuai keahlian masing-masing di bidang perpajakan, pemeriksaan, penilaian, penagihan, dan lainnya sesuai ketentuan peraturan perundang-undangan.

Berdasarkan uraian mengenai struktur organisasi Kantor Pelayanan Pajak (KPP) Pratama Kelompok I dan Kelompok II, dapat disimpulkan bahwa masing-masing kelompok memiliki susunan subbagian dan seksi yang disesuaikan dengan lingkup tugas dan beban kerja di wilayahnya. Setiap bagian dalam organisasi tersebut memiliki peran strategis dalam mendukung pelaksanaan tugas perpajakan, baik dalam hal pelayanan, pengawasan, penjaminan data, hingga penagihan pajak.

Adapun dalam klasifikasi yang telah ditetapkan oleh Direktorat Jenderal Pajak, KPP Pratama Surabaya Gubeng termasuk dalam kategori KPP Pratama Kelompok I. Dengan demikian, struktur organisasi di KPP Pratama Surabaya Gubeng mengikuti susunan organisasi dan pembagian tugas sebagaimana yang telah diuraikan dalam KPP Pratama Kelompok I, yang terdiri dari Subbagian Umum dan Kepatuhan Internal, Seksi Penjaminan Kualitas Data, Seksi Pelayanan, Seksi Pemeriksaan, Penilaian, dan Penagihan, enam Seksi Pengawasan, serta Kelompok Jabatan Fungsional.

### **2.1.3.3. Visi, Misi, dan Tujuan Perusahaan**

#### **2.1.3.3.1. Visi Perusahaan**

Menjadi Mitra Terpercaya Pembangunan Bangsa untuk Menghimpun Penerimaan Negara melalui Penyelenggaraan Administrasi Perpajakan yang

Efisien, Efektif, Berintegritas, dan Berkeadilan dalam rangka mendukung Visi Kementerian Keuangan: "Menjadi Pengelola Keuangan Negara untuk mewujudkan Perekonomian Indonesia yang Produktif, Kompetitif, Inklusif dan Berkeadilan".

#### **2.1.3.3.2. Misi Perusahaan**

1. Merumuskan regulasi perpajakan yang mendukung pertumbuhan ekonomi Indonesia.
2. Meningkatkan kepatuhan pajak melalui pelayanan berkualitas dan terstandarisasi, edukasi dan pengawasan yang efektif, serta penegakan hukum yang adil.
3. Mengembangkan proses bisnis inti berbasis digital didukung budaya organisasi yang adaptif dan kolaboratif serta aparatur pajak yang berintegritas, profesional, dan bermotivasi

Untuk mewujudkan visi dan misinya, Direktorat Jenderal Pajak menyelaraskan tujuan Kementerian Keuangan dengan menetapkan tujuan Direktorat Jenderal Pajak periode 2020 - 2024 yaitu:

1. Pengelolaan fiskal yang sehat dan berkelanjutan.
2. Penerimaan negara yang optimal.
3. Birokrasi dan layanan publik yang agile, efektif, dan efisien.

#### **2.1.3.4. Tugas Pokok dan Fungsi KPP Pratama**

Kantor Pelayanan Pajak (KPP) Pratama merupakan unit pelaksana teknis di lingkungan Direktorat Jenderal Pajak yang bertugas melaksanakan pelayanan, penyuluhan, pengawasan, dan penegakan hukum di bidang perpajakan kepada Wajib Pajak di wilayah kerjanya. Tugas pokok KPP Pratama adalah melaksanakan penerimaan pajak secara optimal melalui



kegiatan administrasi perpajakan yang meliputi pelayanan pendaftaran, penerimaan dan pengolahan Surat Pemberitahuan (SPT), pelayanan permohonan, pelaporan, serta pelaksanaan pemeriksaan, penilaian, penagihan, dan pengawasan kepatuhan perpajakan. Selain itu, KPP Pratama juga bertugas melakukan pembinaan kepada Wajib Pajak agar senantiasa memahami hak dan kewajibannya di bidang perpajakan.

Dalam melaksanakan tugasnya KPP Pratama mempunyai fungsi sebagai berikut :

1. Pengumpulan pencarian dan pengolahan data, pengamatan potensi perpajakan penyajian informasi perpajakan, pendataan objek dan subjek pajak.
2. Penetapan dan penerbitan produk hukum perpajakan.
3. Pengadministrasian dokumen dan berkas perpajakan penerimaan dan pengolahan surat.
4. Pemberitahuan serta penerimaan surat lainnya.
5. Penyuluhan perpajakan.
6. Pelayanan perpajakan.
7. Pelaksanaan pendaftaran wajib pajak.
8. Pelaksanaan Ekstensifikasi.
9. Pengurangan sanksi pajak.
10. Pelaksanaan pemeriksaan pajak.
11. Pengawasan kepatuhan kewajiban perpajakan wajib pajak.
12. Pelaksanaan konsultasi perpajakan.
13. Pembetulan ketetapan pajak.
14. Pelaksanaan administrasi kantor.

#### **2.1.4. Hak dan Wewenang**

##### **2.1.4.1. Hak**

Dalam menjalankan tugas dan fungsinya, KPP Pratama Surabaya Gubeng memiliki sejumlah hak yang bertujuan untuk menunjang kelancaran pelaksanaan administrasi perpajakan. Adapun hak tersebut meliputi:

1. Memperoleh data, dokumen, dan informasi yang diperlukan dalam rangka administrasi perpajakan dari wajib pajak maupun pihak ketiga.
2. Melakukan pemeriksaan terhadap wajib pajak yang terdaftar di wilayah kerjanya.
3. Mengakses dan menggunakan sistem informasi perpajakan untuk mendukung pelaksanaan tugas.
4. Meminta klarifikasi dan penjelasan dari wajib pajak terkait kewajiban dan hak perpajakannya.
5. Melakukan pembetulan ketetapan pajak apabila terdapat kekeliruan sesuai ketentuan.
6. Mendapatkan dukungan data dan informasi dari unit kerja lain di lingkungan Direktorat Jenderal Pajak.

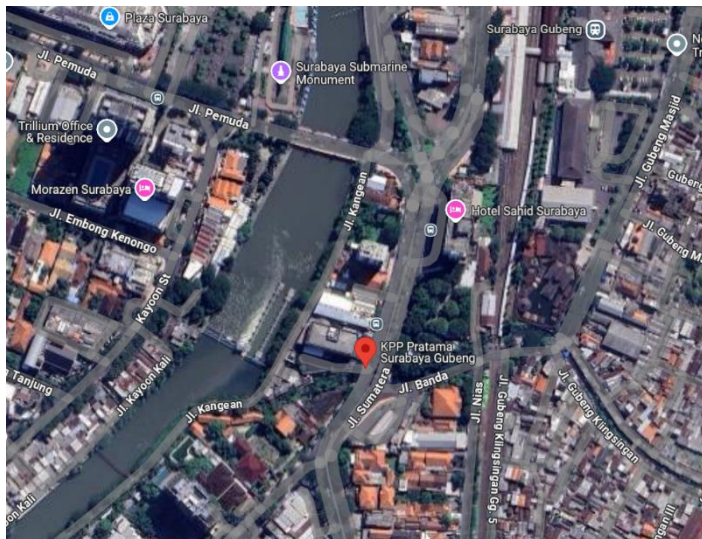
#### **2.1.4.2. Wewenang**

Selain memiliki hak, KPP Pratama Surabaya Gubeng juga diberikan wewenang yang digunakan untuk memastikan kepatuhan dan optimalisasi penerimaan pajak sesuai ketentuan yang berlaku. Wewenang tersebut antara lain:

1. Melaksanakan administrasi penerimaan dan pengawasan pemenuhan kewajiban perpajakan.
2. Melakukan penetapan dan penerbitan produk hukum perpajakan seperti SKP dan STP.
3. Melaksanakan pemeriksaan, penilaian properti dan bisnis, serta tindakan penagihan atas tunggakan pajak.
4. Memberikan pelayanan perpajakan berupa pendaftaran, konsultasi, penyuluhan, dan pengelolaan dokumen.
5. Melakukan intensifikasi dan ekstensifikasi pajak melalui pendataan, pemetaan, dan pengawasan wajib pajak.
6. Menjatuhkan sanksi administratif serta mengurangi atau membatalkan sanksi sesuai peraturan.
7. Mengelola dan menjaga kerahasiaan data serta informasi perpajakan.

### 2.1.5. Lokasi Perusahaan

KPP Pratama Surabaya Gubeng berlokasi di Jl. Ngagel Jaya Selatan No. 93-95, Baratajaya, Kecamatan Gubeng, Kota Surabaya, Jawa Timur. Lokasi ini berada di kawasan strategis yang mudah diakses dari berbagai penjuru kota, baik melalui jalur utama maupun transportasi umum, sehingga memudahkan wajib pajak maupun pihak terkait dalam mendapatkan layanan administrasi perpajakan. Selain itu, keberadaan KPP Pratama Surabaya Gubeng di area perkantoran yang berkembang pesat ini turut mendukung kelancaran operasional serta pelayanan yang optimal kepada masyarakat dan wajib pajak di wilayah kerjanya.



**Gambar 2. 7.** Lokasi KPP Pratama Surabaya Gubeng

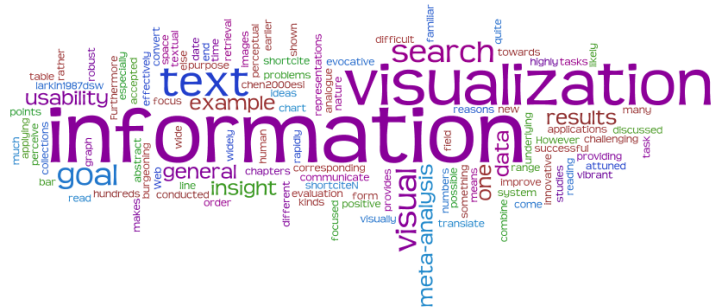
**Sumber:** Google Maps

## 2.2. Landasan Teori

### 2.2.1. Data Teks

Data teks merupakan jenis data tidak terstruktur (*unstructured data*) yang berupa kumpulan kata, kalimat, atau dokumen dalam bentuk digital yang berasal dari berbagai sumber seperti media sosial, ulasan pengguna, artikel berita, email,

ataupun dokumen institusi. Berbeda dengan data terstruktur yang tersimpan dalam format tabel dan database, data teks memerlukan proses khusus untuk diekstraksi dan dianalisis agar informasi di dalamnya dapat dimanfaatkan secara optimal.



**Gambar 2. 8.** Visualisasi Data Teks  
**Sumber:** EWSolutions. (n.d.). *Textual Data Overview*.

Data teks menyimpan informasi yang sangat kaya namun tersembunyi dalam bentuk narasi alami yang tidak dapat langsung diproses oleh komputer. Oleh karena itu, diperlukan teknik khusus dalam bidang *Natural Language Processing* (NLP) dan *Text Mining* untuk mengolah data teks menjadi informasi yang terstruktur, seperti melakukan klasifikasi, klusterisasi, atau analisis sentimen<sup>[1]</sup>.

Salah satu tantangan utama dalam pengolahan data teks adalah keberagaman gaya bahasa, sinonim, ambiguitas makna, hingga *noise* seperti *emoticon*, tanda baca, dan kesalahan pengetikan yang sering muncul dalam teks digital. Oleh sebab itu, tahapan *preprocessing* seperti *tokenization*, *stopword removal*, *stemming*, dan *normalization* menjadi sangat penting untuk menyederhanakan data teks sebelum dilakukan analisis lebih lanjut.

Dalam konteks layanan publik seperti KPP Pratama Surabaya Gubeng, data teks berupa ulasan masyarakat di platform digital seperti Google Maps dapat menjadi sumber informasi

berharga untuk mengetahui kepuasan dan persepsi masyarakat terhadap layanan yang diberikan. Data ini dapat diolah menggunakan algoritma machine learning untuk menghasilkan insight yang dapat mendukung perbaikan layanan berbasis data.

### 2.2.2. Python

Python adalah bahasa pemrograman tingkat tinggi yang bersifat *interpreted*, *open-source*, dan *multi-paradigma*. Bahasa ini pertama kali dikembangkan oleh Guido van Rossum pada akhir tahun 1980-an dan dirilis secara resmi pada tahun 1991. Python dirancang dengan filosofi yang menekankan pada keterbacaan kode, kemudahan penulisan, serta efisiensi pengembangan perangkat lunak. Keunggulan Python terletak pada sintaksnya yang sederhana, modularitasnya yang tinggi, dan dukungan komunitas global yang sangat aktif, sehingga memudahkan para pengembang maupun peneliti dalam mengimplementasikan berbagai metode komputasi, termasuk di bidang *data science*, *machine learning*, dan pengolahan data teks.



**Gambar 2. 9.** Logo Python

**Sumber:** Python Software Foundation. (n.d.). *Python Programming Language Official Website*.

Python telah menjadi salah satu bahasa pemrograman paling populer di dunia data science karena menyediakan ekosistem pustaka (*library*) yang sangat lengkap dan terintegrasi untuk berbagai kebutuhan analisis data, mulai dari *data preprocessing*, *data visualization*, hingga *modeling* dan *evaluation*. Python menawarkan lingkungan pengembangan yang stabil, sintaks yang ekspresif, dan dokumentasi yang luas, sehingga banyak digunakan dalam implementasi algoritma *machine learning* termasuk untuk

tugas klasifikasi sentimen berbasis teks. Keunggulan lain Python terletak pada fleksibilitasnya yang dapat digunakan untuk berbagai keperluan mulai dari *web scraping*, *text preprocessing*, hingga *model deployment*. Python juga mendukung integrasi dengan berbagai database dan sistem web, menjadikannya bahasa yang sangat ideal untuk membangun *pipeline data end-to-end*, dari pengambilan data hingga visualisasi hasil analisis.

Dalam konteks pengolahan data teks dan natural language processing (NLP), Python menyediakan berbagai pustaka populer yang mempermudah pengembangan sistem analisis sentimen berbasis teks. Berbagai pustaka populer atau *library powerful* yang mendukung proses *text preprocessing*, *text mining*, hingga *natural language processing (NLP)*, di antaranya *nlk (Natural Language Toolkit)*, *textblob*, *scikit-learn*, *gensim*, dan *spacy*. Proses seperti *tokenization*, *stopword removal*, *stemming*, *lemmatization*, hingga *vectorization* dapat dilakukan dengan efisien menggunakan *library* tersebut. Selain untuk pengolahan data teks, Python juga sangat andal digunakan untuk pengambilan data secara otomatis dari internet melalui teknik *web scraping*. Proses ini dilakukan dengan memanfaatkan *library* seperti *BeautifulSoup*, *Selenium*, dan *Requests*, yang dapat digunakan untuk mengekstrak data dari halaman web secara terstruktur. Hal ini sangat bermanfaat untuk penelitian yang memerlukan data *real-time* dari media sosial, portal berita, atau platform layanan publik seperti Google Maps, sebagaimana diterapkan dalam laporan kerja praktik ini<sup>[2]</sup>.

Python menjadi salah satu bahasa pemrograman yang paling populer untuk implementasi algoritma *machine learning*, termasuk dalam proyek analisis sentimen. Kelebihan Python terletak pada fleksibilitasnya dalam integrasi dengan berbagai *platform*, kemampuan scripting yang ringan, serta dukungan komunitas yang luas sehingga memudahkan pengembang dalam mencari solusi atas berbagai permasalahan teknis. Python juga memiliki keunggulan dalam hal waktu eksekusi yang efisien dan kemudahan dalam melakukan eksperimen parameter model berkat fitur-fiturnya yang modular dan fleksibel<sup>[3]</sup>.

### 2.2.3. Text Mining



**Gambar 2. 10.** Ilustrasi Proses Text Mining

**Sumber:** DataFlair. (n.d.). *Text Mining: Concepts, Process, Applications & Techniques*. DataFlair.

*Text Mining* adalah proses otomatisasi ekstraksi informasi bernilai dari data berbasis teks yang tidak terstruktur. Proses ini bertujuan untuk menemukan pola, tren, atau pengetahuan tersembunyi yang sebelumnya tidak terlihat dari kumpulan dokumen atau teks. *Text mining* berfokus pada bagaimana memproses teks dalam jumlah besar, mengekstraksi informasi penting, dan menyajikannya dalam bentuk yang terstruktur untuk dianalisis lebih lanjut. *Text mining* didefinisikan sebagai subbidang dari data mining yang secara khusus menangani data berbasis teks. Karena sebagian besar data yang tersedia di internet bersifat tidak terstruktur dalam bentuk teks, maka *text mining* menjadi teknologi yang sangat penting untuk mengolah data tersebut menjadi informasi bermakna. Proses ini memungkinkan berbagai aplikasi, seperti analisis sentimen, klasifikasi dokumen, deteksi topik, hingga prediksi opini masyarakat terhadap suatu isu tertentu.

Dalam praktiknya, proses *text mining* terdiri dari beberapa tahapan penting yang saling berkaitan. Tahap pertama adalah *text collection*, yaitu proses pengumpulan data teks dari berbagai sumber yang dapat berupa media sosial, artikel berita, forum diskusi, atau basis data internal perusahaan. Data yang terkumpul umumnya masih dalam bentuk mentah dan belum siap untuk dianalisis, sehingga perlu melalui tahap *text preprocessing*. Setelah

melalui tahap *preprocessing*, data teks kemudian ditransformasikan ke dalam bentuk representasi numerik melalui proses *text transformation*. Tahapan selanjutnya adalah proses modeling, yaitu penerapan teknik machine learning seperti *Naive Bayes*, *Support Vector Machine (SVM)*, *Decision Tree*, *K-Nearest Neighbor (KNN)*, atau *Neural Network* untuk menemukan pola atau klasifikasi tertentu pada data teks yang telah ditransformasikan. Hasil dari proses *mining* tersebut kemudian dievaluasi menggunakan metrik evaluasi seperti akurasi, *precision*, *recall*, *F1-score*, atau *ROC-AUC*, guna mengukur tingkat keakuratan dan efektivitas model yang dibangun.

Dalam penerapannya, *text mining* telah digunakan di berbagai sektor dan aplikasi. Salah satu aplikasi populer adalah analisis sentimen yang bertujuan untuk mengetahui opini positif atau negatif yang terdapat pada media sosial, ulasan produk, atau data survei pelanggan. Selain itu, *text mining* juga banyak dimanfaatkan dalam klasifikasi dokumen, *information retrieval*, *topic modeling*, *spam detection*, hingga prediksi tren pasar berbasis analisis berita keuangan. *Text mining* memiliki potensi yang besar dalam bidang kesehatan, keuangan, serta sistem pemerintahan digital, karena mampu mengolah data tidak terstruktur dalam jumlah besar menjadi informasi yang strategis. Dengan demikian, *text mining* menjadi salah satu bidang penting dalam pengolahan *big data modern* yang terus berkembang dan memiliki peran strategis di berbagai sektor.

#### **2.2.4. Analisis Sentimen**

Analisis sentimen merupakan cabang dari pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan untuk mengidentifikasi dan mengkategorikan opini atau perasaan yang terkandung dalam suatu teks, baik itu bersifat positif maupun negatif. Teknik ini menjadi semakin penting seiring dengan pertumbuhan eksponensial data teks yang dihasilkan dari berbagai *platform* digital seperti media sosial, forum diskusi, dan ulasan produk. Melalui analisis sentimen, KPP Pratama Surabaya Gubeng dapat memahami persepsi publik terhadap produk, layanan, atau

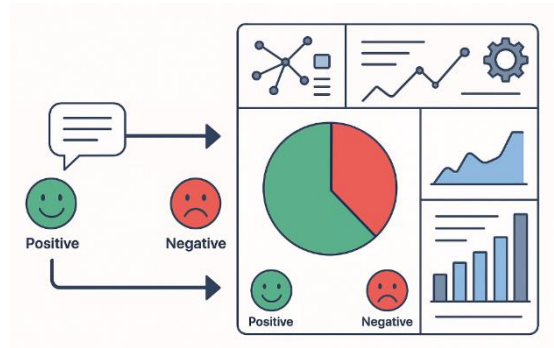


isu tertentu, sehingga memungkinkan pengambilan keputusan yang lebih informasional dan responsif terhadap kebutuhan pengguna secara *real-time*.

Secara umum, pendekatan dalam analisis sentimen dapat dibagi menjadi tiga kategori utama antara lain, berbasis leksikon, pembelajaran mesin (*machine learning*), dan pendekatan hibrida. Pendekatan berbasis leksikon menggunakan daftar kata-kata yang telah diberi label sentimen (positif, negatif) untuk menentukan polaritas teks. Sementara itu, pendekatan *machine learning* melibatkan pelatihan model pada data berlabel untuk mempelajari pola-pola sentimen, menggunakan algoritma seperti *Support Vector Machine* (SVM), *Naive Bayes*, dan berbagai arsitektur *deep learning* seperti *Long Short-Term Memory* (LSTM) dan *Bidirectional Encoder Representations from Transformers* (BERT). Pendekatan hibrida menggabungkan kedua metode tersebut untuk meningkatkan akurasi dan generalisasi model.

Dalam beberapa tahun terakhir, kemajuan signifikan dalam bidang *deep learning* telah mendorong perkembangan analisis sentimen. Model-model seperti BERT dan variannya telah menunjukkan performa yang unggul dalam memahami konteks dan nuansa bahasa alami, memungkinkan analisis sentimen yang lebih akurat, terutama dalam teks yang kompleks atau ambigu. Membandingkan berbagai teknik *deep learning* dalam analisis sentimen, menunjukkan bahwa model berbasis *transformer* seperti BERT memiliki keunggulan dalam menangani konteks yang lebih luas dibandingkan dengan model tradisional seperti CNN dan RNN<sup>[4]</sup>.

Tantangan dalam analisis sentimen juga mencakup penanganan bahasa informal, ironi, sarkasme, dan konteks budaya yang berbeda. Meskipun telah banyak kemajuan, masih terdapat aspek-aspek dalam analisis sentimen yang belum sepenuhnya dieksplorasi, seperti pemahaman emosi yang lebih dalam. Integrasi data dari berbagai modalitas, seperti teks, audio, dan visual, dikenal sebagai analisis sentimen multimodal, yang semakin penting dalam era konten digital yang beragam<sup>[5]</sup>.



**Gambar 2. 11.** Ilustrasi Hasil Analisis Sentimen

**Sumber:** Dokumentasi Pribadi

Dalam konteks aplikasi, analisis sentimen digunakan secara luas dalam berbagai bidang, termasuk pemasaran, layanan pelanggan, politik, dan kesehatan. Kemampuannya untuk memberikan wawasan tentang opini publik menjadikannya alat yang berharga dalam pengambilan keputusan strategis. Namun, penting untuk mempertimbangkan aspek etika dan privasi dalam pengumpulan dan analisis data sentimen, terutama ketika berhadapan dengan data pribadi atau sensitif.

### 2.2.5. Google Maps

Google Maps adalah platform pemetaan digital yang dikembangkan oleh Google, menyediakan berbagai layanan seperti peta jalan, citra satelit, tampilan panorama 360° (*Street View*), informasi lalu lintas waktu nyata, dan perencanaan rute untuk berbagai moda transportasi. Diluncurkan pada tahun 2005, Google Maps telah berevolusi menjadi alat penting dalam kehidupan sehari-hari, digunakan oleh lebih dari satu miliar orang setiap bulannya di seluruh dunia.



## Google Maps

**Gambar 2. 12.** Logo Google Maps

**Sumber:** Vogue. (n.d.). *Google Maps Logo Download (PNG & SVG)*

Secara teknis, Google Maps dibangun menggunakan kombinasi bahasa pemrograman seperti C++ untuk *backend* dan JavaScript, XML, serta Ajax untuk *interface* pengguna. *Platform* ini memungkinkan pengguna untuk berinteraksi dengan peta secara dinamis, seperti memperbesar, memperkecil, dan menavigasi ke berbagai lokasi tanpa perlu memuat ulang halaman. Selain itu, Google Maps menyediakan API (*Application Programming Interface*) yang memungkinkan pengembang untuk mengintegrasikan layanan peta ke dalam aplikasi atau situs web mereka, memperluas fungsionalitas dan aksesibilitas layanan ini<sup>[6]</sup>.

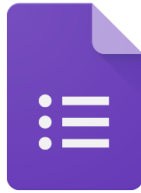
Salah satu fitur unggulan Google Maps adalah kemampuannya dalam memprediksi waktu kedatangan (*Estimated Time of Arrival/ETA*) dengan akurasi tinggi. Untuk mencapai hal ini, Google Maps memanfaatkan teknologi *Graph Neural Networks* (GNN) yang mampu memodelkan interaksi spasial dan temporal dalam jaringan transportasi. Penelitian oleh Derrow-Pinion et al. (2021) menunjukkan bahwa implementasi GNN dalam Google Maps berhasil mengurangi kesalahan prediksi ETA secara signifikan di berbagai wilayah, seperti penurunan lebih dari 40% di kota Sydney<sup>[7]</sup>. Selain itu, Google Maps juga telah mengembangkan fitur-fitur canggih lainnya, seperti *Immersive View*, yang menggabungkan citra *Street View* dan citra udara untuk menghasilkan visualisasi 3D yang realistis dari suatu lokasi. Fitur ini menggunakan teknik pemodelan prediktif dan *neural radiance*

*fields* untuk menciptakan pengalaman navigasi yang lebih imersif dan informatif bagi pengguna.

Dalam konteks pengembangan aplikasi, Google Maps *Platform* menyediakan berbagai API yang dikelompokkan ke dalam tiga kategori utama yaitu, Maps, Places, dan Routes. API ini memungkinkan pengembang untuk mengakses data peta, informasi tempat, dan perencanaan rute, serta mengintegrasikan fitur-fitur seperti geolokasi, pencarian tempat, dan navigasi ke dalam aplikasi mereka. Dengan struktur harga berbasis penggunaan (*pay-as-you-go*), Google Maps *Platform* memberikan fleksibilitas bagi pengembang dalam mengelola biaya dan skala aplikasi mereka. Secara keseluruhan, Google Maps merupakan contoh nyata dari integrasi teknologi geospasial, pembelajaran mesin, dan antarmuka pengguna yang intuitif untuk menyediakan layanan pemetaan yang komprehensif dan mudah diakses. Dengan terus berinovasi dan mengadopsi teknologi terbaru, Google Maps tetap menjadi pemimpin dalam layanan pemetaan digital, memberikan manfaat yang luas bagi individu, bisnis, dan masyarakat secara keseluruhan.

#### **2.2.6. Google Forms**

Google Forms adalah salah satu layanan berbasis web yang dikembangkan oleh Google, yang memungkinkan pengguna untuk membuat dan mendistribusikan formulir survei secara online. Alat ini menyediakan berbagai jenis pertanyaan seperti pilihan ganda, isian singkat, skala linear, dan lainnya, serta mendukung pengumpulan data secara *real-time*. Data yang dikumpulkan secara otomatis disimpan dalam Google Sheets, memudahkan analisis dan visualisasi data.



## Google Forms

**Gambar 2. 13.** Logo Google Forms

**Sumber:** Corkboard Concepts. (2024). *Why Your Business Should Be Using Google Forms.*

Keunggulan Google Forms dalam pengumpulan data kuesioner meliputi aksesibilitas yang tinggi karena dapat diakses melalui berbagai perangkat seperti komputer, tablet, dan smartphone selama terhubung dengan internet. Selain itu, kemudahan penggunaan menjadi salah satu kelebihan karena antarmuka yang intuitif memungkinkan pengguna tanpa latar belakang teknis untuk membuat dan mendistribusikan kuesioner dengan mudah. Google Forms juga memberikan efisiensi waktu dan biaya karena mampu mengurangi kebutuhan akan kertas serta proses manual, sehingga mempercepat proses pengumpulan dan pengolahan data. Keunggulan lainnya yaitu integrasi langsung dengan layanan Google lainnya, seperti Google Sheets untuk analisis data secara otomatis dan kemudahan integrasi dengan aplikasi lain. Tidak hanya itu, Google Forms juga memiliki fitur kolaborasi yang memungkinkan beberapa pengguna untuk bekerja secara bersamaan dalam pembuatan maupun analisis formulir.

Dalam konteks penelitian, Google Forms telah digunakan secara luas untuk mengumpulkan data kuesioner secara online. Misalnya, sebuah penelitian oleh Mubarak et al. (2020) menggunakan model DeLone & McLean untuk menganalisis kepuasan pengguna terhadap layanan Google Forms sebagai media survei online. Hasilnya menunjukkan bahwa kualitas sistem, informasi, dan layanan memiliki pengaruh signifikan terhadap

kepuasan pengguna, menegaskan efektivitas Google Forms dalam pengumpulan data<sup>[8]</sup>.

Google Forms merupakan alat yang efektif dan efisien untuk pengumpulan data melalui kuesioner online. Dengan fitur-fitur yang mendukung kemudahan penggunaan, distribusi yang luas, analisis data otomatis, serta keamanan dan privasi, Google Forms menjadi pilihan yang tepat bagi peneliti, pendidik, dan profesional dalam mengumpulkan dan menganalisis data secara digital

### 2.2.7. *Web Scraping*



**Gambar 2. 14.** Ilustrasi Web Scraping untuk Machine Learning  
**Sumber:** WebScraping.AI. (n.d.). *Web Scraping for Machine Learning*.

*Web scraping*, atau ekstraksi data web, adalah proses otomatisasi pengambilan informasi dari situs web. Teknik ini memungkinkan pengguna untuk mengumpulkan data dalam jumlah besar dari internet secara efisien, yang kemudian dapat digunakan untuk berbagai tujuan seperti analisis data, penelitian, pemantauan harga, dan pengembangan aplikasi. *Web scraping* biasanya melibatkan pengambilan konten HTML dari halaman web dan kemudian mengekstraksi informasi yang relevan menggunakan berbagai metode, termasuk parsing DOM, ekspresi reguler, dan teknik pemrosesan bahasa alami. Dalam praktiknya, *web scraping* dapat dilakukan menggunakan berbagai alat dan bahasa pemrograman. Python, misalnya, menawarkan *library* seperti *BeautifulSoup* dan *Scrapy* yang memudahkan proses pengambilan dan pengolahan data dari web. Selain itu, untuk situs web yang dinamis dan menggunakan JavaScript secara ekstensif, alat seperti Selenium dapat digunakan untuk mensimulasikan

interaksi pengguna dan mengakses konten yang dihasilkan secara dinamis<sup>[9]</sup>.

Teknik *web scraping* melibatkan penggunaan skrip atau program yang dapat mengakses halaman web, mengekstrak informasi yang relevan, dan menyimpannya dalam format yang dapat dianalisis lebih lanjut. Dalam kasus Google Maps, data ulasan biasanya mencakup teks ulasan, rating bintang, tanggal ulasan, dan informasi pengguna. Namun, *web scraping* bukan tanpa tantangan. Salah satu isu utama adalah keberagaman struktur dan format data di berbagai situs web, yang memerlukan pendekatan yang fleksibel dan adaptif dalam proses ekstraksi. Selain itu, beberapa situs web menerapkan mekanisme *anti-scraping* seperti CAPTCHA, pembatasan IP, dan deteksi bot, yang dapat menghambat proses pengambilan data. Untuk mengatasi tantangan ini, penelitian terbaru telah mengembangkan model seperti *WebFormer*, yang menggunakan pendekatan transformer untuk mengekstraksi informasi terstruktur dari dokumen web dengan mempertimbangkan tata letak dan struktur HTML.

Dalam penelitian oleh Brenning dan Henn (2023), *web scraping* diidentifikasi sebagai alat yang menjanjikan untuk akuisisi data geografis. Mereka menyoroti bahwa *web scraping* memungkinkan akses hampir *real-time* ke data geolokasi objek, yang dapat diperoleh dengan biaya efektif. Namun, mereka juga mencatat tantangan unik yang dihadapi, termasuk isu etika dan legal terkait hak kekayaan intelektual, privasi, dan integritas situs web. Selain itu, tantangan teknis dan statistik seperti ketergantungan, ketidaklengkapan data, inkonsistensi, dan bias juga menjadi perhatian utama dalam praktik *web scraping*<sup>[10]</sup>.

Dalam studi lain oleh Alaydaa et al. (2023), analisis sentimen berbasis aspek diterapkan pada ulasan pelancong yang dikumpulkan dari Google Maps tentang bandara Dubai dan Doha. Penelitian ini menunjukkan bahwa dengan menggunakan *web scraping* untuk mengumpulkan data ulasan, analisis yang lebih mendalam dan spesifik terhadap aspek-aspek layanan bandara dapat dilakukan, memberikan wawasan yang lebih tajam untuk peningkatan kualitas layanan<sup>[11]</sup>.

Secara keseluruhan, *web scraping* merupakan teknik yang kuat untuk mengumpulkan data ulasan dari Google Maps, yang dapat digunakan untuk berbagai analisis, termasuk analisis sentimen dan evaluasi layanan. Namun, penting untuk selalu mempertimbangkan aspek legal, etika, dan teknis dalam pelaksanaannya, serta memastikan bahwa data yang dikumpulkan digunakan secara bertanggung jawab dan sesuai dengan peraturan yang berlaku.

### 2.2.8. Preprocessing Data

*Preprocessing data* merupakan tahap krusial dalam proses analisis data teks, terutama dalam domain *Natural Language Processing* (NLP) dan analisis sentimen. Tahap ini bertujuan untuk membersihkan dan mempersiapkan data mentah agar dapat diproses lebih lanjut oleh algoritma pembelajaran mesin atau model statistik. Dalam konteks ulasan pengguna di platform seperti Google Maps, data yang diperoleh sering kali tidak terstruktur, mengandung kesalahan penulisan, penggunaan bahasa informal, serta elemen-elemen non-teks seperti emoji atau simbol. Oleh karena itu, preprocessing menjadi langkah awal yang esensial untuk meningkatkan kualitas data dan akurasi analisis.



**Gambar 2. 15.** Ilustrasi Teknik-Teknik Text Preprocessing

**Sumber:** Somiya. (2021). *Understanding Fundamental Text Preprocessing Techniques for NLP*. Medium.



Langkah-langkah umum dalam preprocessing data teks meliputi:

1. Pembersihan Teks (*Text Cleaning*): Menghapus karakter-karakter yang tidak diperlukan seperti angka, tanda baca, dan simbol-simbol khusus yang tidak memberikan informasi signifikan terhadap analisis.
2. *Case Folding*: Mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menyamakan format, sehingga kata seperti "Data" dan "data" dianggap sama.
3. Normalisasi Teks: Mengubah seluruh teks menjadi format yang konsisten, dengan memperbaiki kesalahan penulisan atau penggunaan kata-kata slang.
4. Tokenisasi (*Tokenizing*): Memecah teks menjadi unit-unit kata atau frasa yang lebih kecil (*tokens*) untuk memudahkan analisis lebih lanjut.
5. Penghapusan Kata Umum (*Stopword Removal*): Menghapus kata-kata yang sering muncul namun tidak memiliki makna signifikan dalam analisis, seperti "dan", "yang", atau "adalah".
6. *Stemming*: Mengubah kata-kata ke bentuk dasarnya untuk menyatukan variasi kata yang memiliki makna serupa.

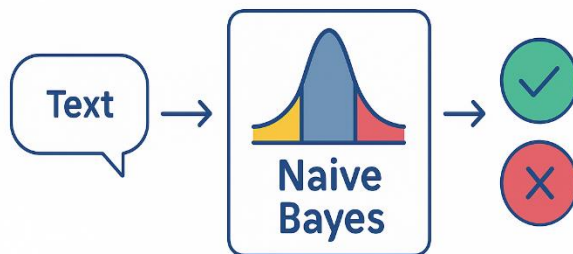
Dalam studi oleh Aufar et al. (2023), diterapkan berbagai teknik *preprocessing* pada ulasan aplikasi *e-wallet* DANA yang diambil dari Google Play Store. Hasil penelitian menunjukkan bahwa kombinasi teknik seperti pembersihan teks, *case folding*, tokenisasi, dan lemmatisasi mampu meningkatkan akurasi analisis sentimen secara signifikan, dengan rata-rata akurasi mencapai 99,63% menggunakan model *Support Vector Machine* (SVM) dengan berbagai kernel<sup>[12]</sup>. Penelitian lebih lanjut, Arief dan Deris (2021) dalam konferensi Internasional ICIC meneliti dampak preprocessing terhadap klasifikasi sentimen ulasan produk. Mereka menemukan bahwa *preprocessing* yang efektif dapat meningkatkan kinerja model klasifikasi seperti *Decision Tree*, *Naïve Bayes*, dan SVM, dengan SVM menunjukkan akurasi tertinggi sebesar 88,13%<sup>[13]</sup>.

Dalam konteks ulasan Google Maps, tantangan tambahan muncul karena data ulasan dapat mencakup berbagai bahasa,

dialek, serta penggunaan bahasa informal atau lokal. Oleh karena itu, preprocessing harus disesuaikan dengan karakteristik data tersebut, termasuk penanganan bahasa campuran, transliterasi, serta pengenalan dan penghapusan entitas khusus seperti nama tempat atau produk.

Secara keseluruhan, *preprocessing data* merupakan langkah fundamental yang menentukan keberhasilan analisis sentimen, terutama ketika berhadapan dengan data ulasan dari platform seperti Google Maps. Dengan menerapkan teknik *preprocessing* yang tepat dan disesuaikan dengan karakteristik data, analisis sentimen dapat menghasilkan informasi yang lebih akurat dan bermanfaat untuk pengambilan keputusan strategis.

### 2.2.9. Naive Bayes



**Gambar 2. 16.** Diagram Konsep Algoritma Naive Bayes

**Sumber:** Dokumentasi Pribadi

*Naive Bayes* adalah algoritma klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur. Meskipun asumsi ini sering kali tidak realistis dalam praktik, pendekatan ini tetap efektif dalam berbagai aplikasi, termasuk analisis sentimen. Dalam konteks analisis sentimen terhadap ulasan pengguna di Google Maps, *Naive Bayes* digunakan untuk mengklasifikasikan teks ulasan menjadi kategori sentimen seperti positif, negatif, atau netral. Model ini menghitung

probabilitas suatu teks termasuk dalam kategori tertentu berdasarkan frekuensi kata-kata yang muncul dalam teks tersebut.

Tujuan utama dari *Naive Bayes Classifier* adalah menghitung probabilitas posterior dari suatu kelas  $C_k$  berdasarkan fitur-fitur input  $x_1, x_2, \dots, x_n$  dengan menggunakan Teorema Bayes, probabilitas ini dapat dinyatakan sebagai:

$$P(C_k | x_1, x_2, \dots, x_n) = \frac{P(C_k) \cdot P(x_1, x_2, \dots, x_n | C_k)}{P(x_1, x_2, \dots, x_n)}$$

Karena  $P(x_1, x_2, \dots, x_n)$  bersifat konstan untuk semua kelas, fokus utama adalah pada bagian numerator. Dengan asumsi independensi antar fitur,  $P(x_1, x_2, \dots, x_n | C_k)$  dapat diuraikan menjadi:

$$P(C_k | x_1, x_2, \dots, x_n) \propto P(C_k) \cdot \prod_{i=1}^n P(x_i | C_k)$$

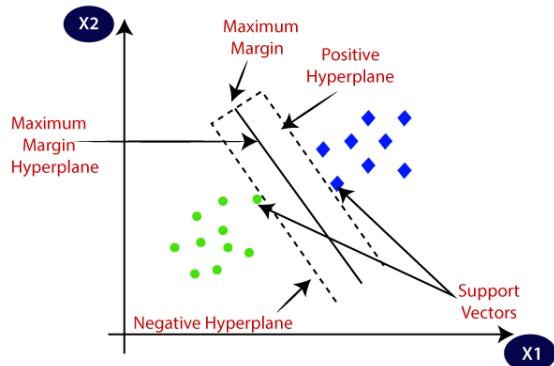
Klasifikasi dilakukan dengan memilih kelas  $C_k$  yang memaksimalkan probabilitas posterior ini, atau dimana:

1.  $P(C_k | x_1, x_2, \dots, x_n)$  adalah probabilitas posterior dari kelas  $C_k$  diberikan fitur-fitur  $x_1, x_2, \dots, x_n$ .
2.  $P(C_k)$  adalah probabilitas prior dari kelas  $C_k$ .
3.  $P(x_i | C_k)$  adalah probabilitas likelihood dari fitur  $x_i$  diberikan kelas  $C_k$ .

Dalam analisis sentimen terhadap ulasan pengguna di Google Maps, *Naive Bayes* dapat digunakan untuk mengklasifikasikan ulasan berdasarkan sentimen yang terkandung di dalamnya. Keunggulan utama dari *Naive Bayes* dalam konteks ini adalah kemampuannya untuk menangani dataset besar dengan efisiensi tinggi dan kebutuhan komputasi yang relatif rendah. Namun, asumsi independensi antar fitur dapat menjadi keterbatasan,

terutama ketika terdapat korelasi kuat antar kata dalam teks ulasan<sup>[14]</sup>.

### 2.2.10. Support Vector Machine (SVM)



**Gambar 2. 17.** Ilustrasi Konsep Support Vector Machine (SVM)  
**Sumber:** Datasciencelovers.com (2023)

*Support Vector Machine* (SVM) adalah salah satu algoritma dalam bidang pembelajaran mesin (*machine learning*) yang banyak digunakan untuk tugas-tugas klasifikasi dan regresi. SVM bekerja dengan cara mencari sebuah *hyperplane* atau bidang pemisah terbaik di dalam ruang berdimensi banyak yang dapat memisahkan data ke dalam dua atau lebih kelas dengan margin pemisahan maksimal. *Margin* ini merupakan jarak antara *hyperplane* dengan titik data terdekat dari masing-masing kelas. Semakin besar *margin* yang dihasilkan, semakin baik kemampuan model dalam membedakan kelas-kelas tersebut, karena dapat meningkatkan generalisasi saat menghadapi data baru.

*Hyperplane* dapat dinyatakan dengan persamaan:

$$w^T x + b = 0$$

Di mana:

1.  $w$  adalah vektor bobot.
2.  $x$  adalah vektor fitur input.
3.  $b$  adalah bias.

Untuk memastikan bahwa data diklasifikasikan dengan benar dan margin maksimal, SVM memecahkan masalah optimasi berikut:

$$\begin{aligned} & \underset{\mathbf{w}, b}{\text{minimize}} && \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{subject to} && y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \quad \forall i \end{aligned}$$

Di mana  $y_i$  adalah label kelas dari data ke- $i$ , yang bernilai +1 atau -1.

Dalam implementasinya, SVM tidak hanya bekerja pada data yang bersifat linear (dapat dipisahkan secara langsung oleh sebuah garis lurus atau bidang datar), tetapi juga sangat efektif untuk data *non-linear*. Hal ini dimungkinkan melalui penggunaan teknik yang disebut *kernel trick*, yaitu metode yang memetakan data ke dalam dimensi yang lebih tinggi agar data tersebut dapat dipisahkan secara linear. Beberapa jenis kernel yang umum digunakan dalam SVM antara lain *linear kernel*, *polynomial kernel*, *radial basis function (RBF) kernel*, dan *sigmoid kernel*. Pemilihan kernel yang tepat sangat berpengaruh terhadap performa model SVM dalam menyelesaikan permasalahan klasifikasi maupun regresi.

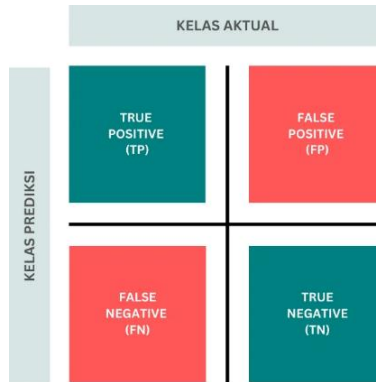
Beberapa fungsi *kernel* yang umum digunakan antara lain:

1. *Linear Kernel*:  $K(x_i, x_j) = x_i^\top x_j$
2. *Polynomial Kernel*:  $K(x_i, x_j) = (\gamma x_i^\top x_j + r)^d$
3. *Radial Basis Function (RBF) Kernel*:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$
4. *Sigmoid Kernel*:  $K(x_i, x_j) = \tanh(\alpha \cdot x_i^\top x_j + c)$

Salah satu aplikasi penting dari SVM adalah dalam bidang *text mining* dan *analisis sentimen*. Pada konteks ini, SVM digunakan untuk mengklasifikasikan teks ke dalam kategori sentimen tertentu, misalnya positif atau negatif. Prosesnya dimulai dengan melakukan ekstraksi fitur dari data teks, seperti menggunakan metode *Bag of Words* (BoW), TF-IDF (*Term Frequency - Inverse Document Frequency*), atau *word embeddings* untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma SVM. Selanjutnya, model SVM dilatih menggunakan data yang telah diberi label sentimen, sehingga model mampu mempelajari pola-pola tertentu dari teks yang berkaitan dengan sentimen yang dimaksud.

#### **2.2.11. Evaluasi Model Klasifikasi**

Evaluasi model klasifikasi merupakan tahap krusial dalam proses pengembangan sistem pembelajaran mesin, khususnya dalam konteks analisis sentimen. Tujuan utama dari evaluasi ini adalah untuk menilai sejauh mana model mampu mengklasifikasikan data ke dalam kategori sentimen yang tepat, seperti positif atau negatif. Proses evaluasi ini tidak hanya membantu dalam menilai kinerja model, tetapi juga dalam mengidentifikasi area yang memerlukan perbaikan untuk meningkatkan akurasi dan efisiensi sistem. Dalam analisis sentimen, berbagai metrik digunakan untuk mengevaluasi performa model, antara lain akurasi, presisi, *recall*, dan *F1-score*.



**Gambar 2. 18.** Diagram Confusion Matrix  
**Sumber:** Sainsdata.id (2023)

Beberapa metrik umum yang digunakan dalam evaluasi model klasifikasi meliputi:

1. Akurasi (*Accuracy*)

Akurasi mengukur proporsi prediksi yang benar terhadap total prediksi yang dilakukan. Meskipun mudah dipahami, akurasi bisa menyesatkan jika data tidak seimbang.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Di mana:

- a.  $TP$  : True Positive
- b.  $TN$  : True Negative
- c.  $FP$  : False Positive
- d.  $FN$  : False Negative

## 2. Presisi (*Precision*)

Presisi mengukur proporsi prediksi positif yang benar terhadap semua prediksi positif yang dibuat oleh model. Presisi tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang salah.

$$Presisi = \frac{TP}{TP+FP}$$

## 3. Recall (Sensitivitas)

*Recall* mengukur proporsi kasus positif yang benar-benar teridentifikasi oleh model. *Recall* tinggi menunjukkan bahwa model berhasil menangkap sebagian besar kasus positif.

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

## 4. F1-Score

F1-Score adalah rata-rata harmonis dari presisi dan *recall*, memberikan keseimbangan antara keduanya. *F1-Score* berguna ketika terdapat ketidakseimbangan antara presisi dan *recall*.

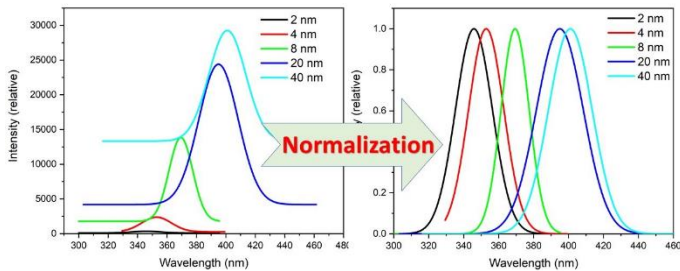
$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

*F1-Score* sangat relevan dalam aplikasi seperti analisis sentimen, di mana penting untuk menyeimbangkan antara mengidentifikasi semua ulasan positif dan memastikan bahwa ulasan yang diklasifikasikan sebagai positif memang benar-benar positif<sup>[15]</sup>.

Evaluasi model klasifikasi adalah langkah krusial dalam pengembangan sistem analisis sentimen yang andal, karena dapat membantu dalam menilai kinerja model secara menyeluruh, terutama dalam situasi dengan data yang tidak seimbang.



## 2.2.12. Uji Normalisasi Data



**Gambar 2. 19.** Transformasi Uji Normalisasi data

**Sumber:** YouTube, “Machine Learning NLP Text Preprocessing Tutorial”

Dalam proses analisis data, khususnya pada penelitian yang melibatkan pemrosesan data ulasan teks seperti di Google Maps, tahap *preprocessing data* menjadi aspek yang sangat penting untuk memastikan kualitas data yang optimal sebelum diterapkan ke dalam model pembelajaran mesin. Salah satu tahapan dalam proses ini adalah uji normalisasi data atau *Data Normalization Test*, yaitu proses untuk menilai apakah data telah memenuhi asumsi distribusi tertentu, khususnya distribusi normal dan untuk mengubah skala nilai fitur numerik dalam dataset agar berada dalam rentang tertentu dan memiliki bobot kontribusi yang setara saat diproses oleh algoritma klasifikasi. Normalisasi menjadi krusial jika hasil performanya belum optimal, karena algoritma seperti *Support Vector Machine* (SVM) dan *Naive Bayes* cenderung sensitif terhadap perbedaan skala antar fitur, yang dapat menyebabkan bias dalam proses pembelajaran model.

Metode normalisasi yang apabila digunakan dalam penelitian ini adalah *Min-Max Scaling*, yaitu metode yang mentransformasi nilai data ke dalam rentang [0,1], dengan rumus:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Di mana  $X_{norm}$  adalah nilai hasil normalisasi,  $X$  adalah nilai asli data,  $X_{min}$  adalah nilai minimum dari fitur tersebut, dan  $X_{max}$  adalah nilai maksimumnya. Dengan menggunakan *Min-Max Scaling*, semua nilai fitur dinyatakan dalam rentang 0 hingga 1, tanpa mengubah bentuk distribusi asli datanya. Penerapan metode ini sangat bermanfaat pada analisis sentimen berbasis teks karena seringkali nilai-nilai fitur yang dihasilkan dari proses ekstraksi teks seperti frekuensi kata, skor sentimen, atau bobot *TF-IDF* memiliki skala yang berbeda-beda. Apabila tidak dinormalisasi, fitur dengan nilai yang besar dapat mendominasi hasil model, sehingga mempengaruhi akurasi prediksi.

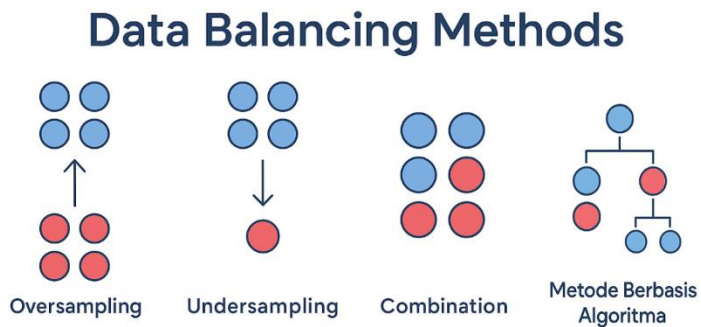
Uji normalisasi data dengan *Min-Max Scaling* secara signifikan meningkatkan kinerja klasifikasi sentimen teks pada berbagai *platform* media sosial. Penelitian tersebut membuktikan bahwa normalisasi berbasis rentang efektif dalam meningkatkan stabilitas nilai parameter model klasifikasi serta mempercepat proses konvergensi algoritma pembelajaran mesin, terutama dalam menangani data ulasan yang memiliki nilai variabel yang beragam. Selain itu, normalisasi ini juga mempermudah proses interpretasi nilai parameter model, karena semua fitur sudah berada dalam skala yang seragam. Oleh sebab itu, dalam konteks penelitian analisis sentimen terhadap ulasan Google Maps, penerapan *Min-Max Scaling* dapat dianggap sebagai langkah strategis untuk meningkatkan keandalan model klasifikasi yang digunakan, baik untuk algoritma *Naive Bayes* maupun *Support Vector Machine*, dalam mengelompokkan opini positif maupun negatif dari Wajib Pajak KPP Pratama Surabaya Gubeng<sup>[16]</sup>.

### 2.2.13. *Data Balancing Methods*

Pada tugas klasifikasi seperti analisis sentimen, seringkali dihadapi masalah ketidakseimbangan kelas (*class imbalance*), di mana jumlah sampel dalam satu kelas (misalnya, ulasan positif) jauh lebih banyak dibandingkan kelas lainnya (misalnya, ulasan negatif). Ketidakseimbangan ini dapat menyebabkan model pembelajaran mesin cenderung memprediksi kelas mayoritas,

sehingga mengabaikan kelas minoritas yang mungkin lebih penting dalam konteks tertentu.

Data Balancing Methods adalah proses untuk mengidentifikasi dan mengatasi ketidakseimbangan kelas dalam dataset. Tujuannya adalah untuk memastikan bahwa model pembelajaran mesin dapat belajar secara efektif dari semua kelas yang ada, tanpa bias terhadap kelas mayoritas. Ketidakseimbangan distribusi kelas sering menjadi tantangan yang dapat memengaruhi kinerja model prediksi, terutama dalam klasifikasi. Oleh karena itu, penting untuk menerapkan teknik khusus yang dirancang untuk menangani permasalahan ini agar hasil analisis tetap akurat dan representatif.



**Gambar 2. 20.** Transformasi Uji Normalisasi data

**Sumber:** Dokumentasi Pribadi (Hasil Desain Menggunakan AI ChatGPT, 2025).

Beberapa pendekatan yang umum diterapkan untuk mengatasi ketidakseimbangan data tersebut di antaranya meliputi:

1. *Oversampling* merupakan metode dengan menambahkan sampel ke kelas minoritas untuk menyamakan jumlahnya dengan kelas mayoritas.
  - a. *Random Oversampling* adalah menggandakan sampel yang ada secara acak.
  - b. *SMOTE (Synthetic Minority Over-sampling Technique)* adalah membuat sampel sintetis baru dengan menginterpolasi antara sampel minoritas yang ada.

2. *Undersampling* merupakan metode dengan mengurangi jumlah sampel di kelas mayoritas.
  - a. *Random Undersampling* adalah menghapus sampel dari kelas mayoritas secara acak.
  - b. *Tomek Links* adalah menghapus sampel yang berada di perbatasan antara kelas mayoritas dan minoritas untuk meningkatkan kejelasan klasifikasi.
3. Metode Gabungan merupakan metode dengan menggabungkan *oversampling* dan *undersampling* untuk mencapai keseimbangan yang optimal.
4. Metode Berbasis Algoritma merupakan metode dengan menggunakan algoritma pembelajaran mesin yang dirancang untuk menangani ketidakseimbangan data, seperti *ensemble learning* atau *cost-sensitive learning*<sup>[17]</sup>.

Dalam konteks analisis sentimen terhadap ulasan pengguna di Google Maps KPP Pratama Surabaya Gubeng, ketidakseimbangan kelas dapat terjadi jika, misalnya, sebagian besar ulasan bersifat positif, sementara ulasan negatif sangat sedikit. Hal ini dapat menyebabkan model klasifikasi tidak mampu mengenali ulasan negatif dengan baik. Dengan menerapkan teknik *undersampling*, seperti *Random Undersampling* atau *Tomek Links*, jumlah sampel dari kelas mayoritas (ulasan positif) dikurangi untuk menyamakan jumlahnya dengan kelas minoritas (ulasan negatif). Hal ini membantu model untuk belajar secara seimbang dari kedua kelas, meningkatkan kemampuannya dalam mengklasifikasikan sentimen secara akurat.

## **BAB 3**

### **METODE PENELITIAN**

#### **3.1. Metode Pengumpulan Data**

Dalam penyusunan laporan kerja praktik ini, penulis memperoleh data dan referensi dengan memanfaatkan beberapa metode pengumpulan data yang dilakukan secara sistematis. Adapun metode yang digunakan adalah sebagai berikut:

##### **3.1.1. Metode Kepustakaan**

Metode ini dilakukan dengan cara mengumpulkan informasi dan teori-teori yang relevan dari berbagai sumber seperti buku, jurnal ilmiah, artikel, laporan penelitian sebelumnya, serta referensi daring yang berkaitan dengan topik klasifikasi sentimen, algoritma *Naive Bayes*, *Support Vector Machine* (SVM), serta konsep pengolahan data teks. Metode ini bertujuan untuk memperoleh landasan teori dan pemahaman yang mendukung pelaksanaan Kerja Praktik.

##### **3.1.2. Metode Web Scraping**

Metode ini digunakan untuk mengumpulkan data sentimen masyarakat berupa ulasan atau komentar yang dipublikasikan di platform Google Maps pada halaman profil KPP Pratama Surabaya Gubeng. Teknik Web Scraping merupakan proses otomatisasi pengambilan data dari halaman web menggunakan program atau skrip tertentu, tanpa harus melakukan pengambilan data secara manual. Pada praktiknya, proses pengambilan data dilakukan dengan memanfaatkan library Python seperti selenium untuk mengekstrak teks ulasan yang nantinya akan dijadikan dataset dalam proses klasifikasi sentimen. Data yang diperoleh melalui metode ini berupa isi *Username*, *Date*, *Rating*, dan *Review* sebagai informasi tambahan untuk analisis sentimen.

##### **3.1.3. Metode Wawancara**

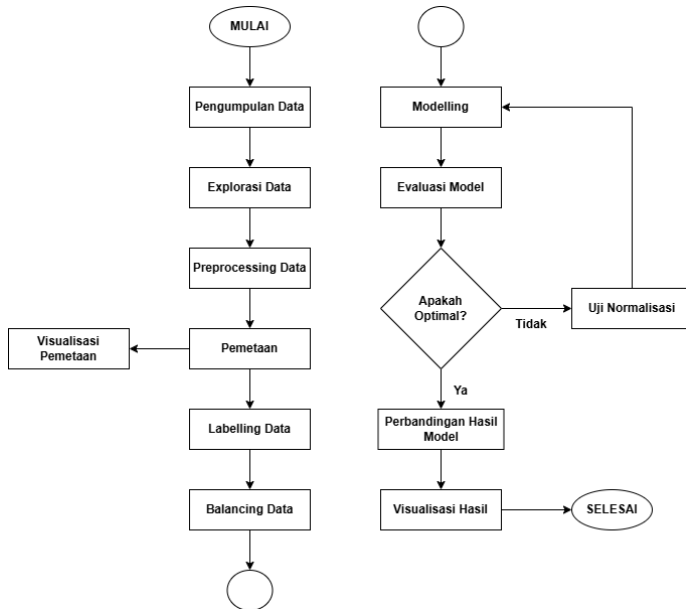
Metode wawancara dilakukan secara langsung dengan pihak terkait di lingkungan KPP Pratama Surabaya Gubeng,

khhususnya pegawai pada bagian subbagian umum dan kepatuhan internal dan pengawasan. Wawancara ini bertujuan untuk memperoleh informasi lebih mendalam mengenai gambaran umum perusahaan, alur pelayanan, jenis layanan yang sering diakses masyarakat, serta kendala yang dihadapi dalam pelayanan perpajakan. Informasi ini menjadi pelengkap dalam memahami konteks data sentimen masyarakat yang dianalisis, serta memastikan relevansi antara ulasan masyarakat di Google Maps dengan kondisi layanan yang ada di lapangan.

#### **3.1.4. Metode Kuesioner**

Metode kuesioner dilakukan dengan menyebarkan daftar pertanyaan menggunakan *platform* Google Forms secara daring melalui media sosial Twitter atau yang sekarang dikenal dengan X kepada responden yang merupakan wajib pajak atau masyarakat pengguna layanan di KPP Pratama Surabaya Gubeng. Tujuan dari metode ini adalah untuk mengumpulkan data ulasan secara *real* mengenai tingkat kepuasan, pengalaman, dan pendapat masyarakat terhadap layanan perpajakan yang diterima serta fasilitas pendukung di lingkungan kantor. Hasil dari kuesioner ini digunakan sebagai data tambahan untuk mendukung analisis sentimen masyarakat dari ulasan Google Maps, serta untuk membandingkan persepsi masyarakat secara langsung dengan ulasan yang tersedia di media digital.

### 3.2. Prosedur Penelitian



**Gambar 3. 1.** Flowchart Prosedur Penelitian

**Sumber:** Dokumentasi Pribadi

Pada subbab ini akan dijelaskan prosedur penelitian yang dilakukan secara sistematis untuk menganalisis sentimen masyarakat terhadap Kantor Pelayanan Pajak Surabaya Gubeng. Prosedur diawali dengan tahap pengumpulan data, yang dilakukan melalui dua metode, yaitu *web scraping* ulasan pengguna di *platform* Google Maps pada akun resmi KPP Pratama Surabaya Gubeng serta menggunakan Google Forms kepada wajib pajak melalui media sosial X untuk menjadi objek penelitian. Selanjutnya dilakukan eksplorasi data untuk memahami karakteristik awal data yang diperoleh. Setelah data berhasil dikumpulkan, tahap selanjutnya adalah *preprocessing data*. Pada tahap ini, data teks dibersihkan dari karakter yang tidak diperlukan (*cleaning*), diubah menjadi huruf kecil (*case folding*), dilakukan normalisasi untuk mengubah kata tidak baku menjadi baku, memecah kalimat menjadi kata-kata tunggal (*tokenizing*), menghapus

kata-kata umum yang tidak memiliki makna penting (*stopword removal*), serta melakukan *stemming* untuk mengubah kata ke bentuk dasar.

Setelah itu, dilakukan pemetaan data, yaitu mengelompokkan isi ulasan berdasarkan kategori layanan, tempat, dan secara keseluruhan (*Overall*) yang disebutkan dalam ulasan. Visualisasi pemetaan kemudian dilakukan untuk melihat distribusi topik ulasan yang muncul. Hasil data ulasan yang telah dipetakan kemudian masuk ke tahapan labelling atau analisis sentimen, yang diklasifikasikan ke dalam dua kategori, yaitu positif dan negatif. Dilanjutkan dengan *balancing data* menggunakan teknik *undersampling* jika terdapat ketidakseimbangan jumlah data antar kategori sentimen. Setelah data seimbang dilakukan proses ekstraksi fitur menggunakan metode TF-IDF untuk mengubah data teks menjadi nilai numerik. Tahap berikutnya adalah *modelling*, di mana data yang telah diekstrak fiturnya akan diklasifikasikan menggunakan dua algoritma *machine learning*, yaitu *Naive Bayes* dan *Support Vector Machine* (SVM). Hasil klasifikasi dari masing-masing model kemudian dievaluasi menggunakan beberapa metrik, yakni akurasi, presisi, *recall*, dan *F1-score*.

Apabila hasil evaluasi menunjukkan performa model yang kurang optimal maka dilakukan uji normalisasi data untuk memastikan distribusi data telah sesuai sebelum proses pelatihan model diulang kembali. Setelah semua tahapan tersebut, dilakukan perbandingan visualisasi hasil evaluasi dari kedua model untuk mengetahui algoritma mana yang memberikan performa terbaik. Seluruh prosedur ini diakhiri dengan penyusunan kesimpulan berdasarkan hasil analisis yang telah dilakukan.

### 3.3. Kontribusi

Pada pelaksanaan Kerja Praktik (KP) di KPP Pratama Surabaya Gubeng, penulis berkesempatan untuk terlibat dalam beberapa kegiatan yang mendukung operasional, pelayanan, hingga pengolahan dan analisis data perpajakan. Adapun kontribusi yang diberikan penulis selama KP antara lain:



### **3.3.1. Membantu Tugas di Subbagian Umum dan Kepatuhan**

Penulis turut berperan dalam membantu pelaksanaan tugas di Subbagian Umum dan Kepatuhan Internal. Kegiatan yang dilakukan meliputi pengelolaan administrasi umum, seperti melakukan pengarsipan surat masuk dan surat keluar secara teratur, mendistribusikan dokumen-dokumen penting kepada unit terkait, serta membantu dalam pelaksanaan berbagai tugas harian yang mendukung kelancaran operasional di lingkungan kantor. Selain itu, penulis juga berkesempatan untuk memahami alur kerja administrasi di lingkungan KPP Pratama Surabaya Gubeng secara langsung.

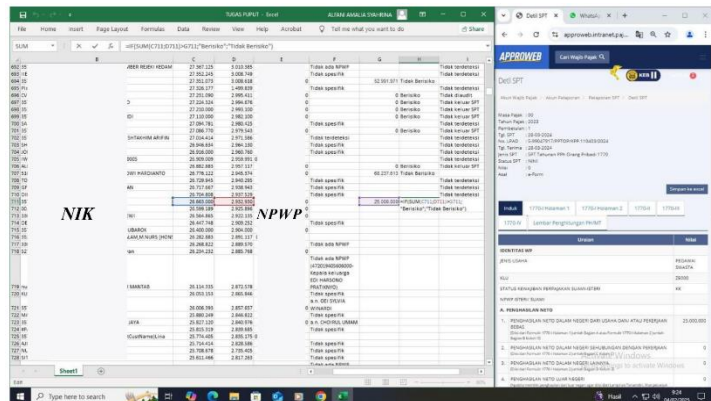
### **3.3.2. Mempelajari Cara Pelayanan Pembuatan SPT**

Penulis diberi kesempatan untuk mempelajari prosedur layanan pembuatan Surat Pemberitahuan Tahunan (SPT) baik untuk Wajib Pajak Pribadi maupun Badan. Kegiatan ini meliputi pendampingan wajib pajak saat melakukan pelaporan SPT secara langsung maupun melalui sistem *e-Filing*, serta memahami alur pengecekan kelengkapan data dan syarat administratif.

### **3.3.3. Blast WhatsApp Reminder SPT**

Turut membantu tugas dari seksi pengawasan VI untuk melakukan *blast* pesan WhatsApp secara massal kepada Wajib Pajak Pribadi dan Badan sebagai pengingat untuk segera melakukan pelaporan SPT Tahunan. Kegiatan ini dilakukan secara manual tanpa menggunakan *tools*, namun data Wajib Pajak dan *template* pesan sudah tersimpan dalam file *Excel*. Proses ini bertujuan meningkatkan kepatuhan wajib pajak dalam memenuhi kewajiban perpajakannya tepat waktu.

### 3.3.4. Analisis Prediksi Risiko Wajib Pajak



Gambar 3. 2. Prediksi Risiko Wajib Pajak

Sumber: Dokumentasi Pribadi

Penulis terlibat dalam kegiatan analisis dan membuat prediksi risiko bagi setiap Wajib Pajak yang data dan NPWP-nya sudah tervalidasi benar adanya oleh *website APPROWEB* milik Direktorat Jenderal Pajak (DJP). Proses penentuan kategori risiko dilakukan dengan menggunakan rumus perhitungan di *Microsoft Excel*, yang membandingkan nilai total Dasar Pengenaan Pajak (DPP) ditambah Pajak Pertambahan Nilai (PPN) dengan total penghasilan Wajib Pajak. Apabila hasil penjumlahan DPP dan PPN melebihi jumlah penghasilan Wajib Pajak, maka Wajib Pajak tersebut diklasifikasikan ke dalam kategori "Berisiko." Sebaliknya, apabila total DPP dan PPN lebih kecil atau sama dengan total penghasilan, maka Wajib Pajak tersebut dikategorikan sebagai "Tidak Berisiko."

Rumus yang digunakan pada *Excel* dalam pengerjaan ini adalah:  

$$=IF(SUM(C8;D8) > G8; Berisiko; Tidak Berisiko)$$

### 3.3.5. Fuzzy Matching Data Restaurant Kecamatan Sukolilo

```
1 import pandas as pd
2 from thefuzz import fuzz
3 from thefuzz import process
4
5 # Load dua data file CSV
6 file1 = pd.read_csv("Restaurant_Sukolilo.csv")
7 file2 = pd.read_csv("UTK_TUGAS_RUPUT_MASKING.csv")
8
9 kolom_alamat1 = "Alamat"
10 kolom_alamat2 = "ALAMAT"
11 kolom_npwp = "IDENTITAS DATA"
12
13 # Fungsi untuk melakukan fuzzy matching
14 def fuzz_match(row, daftar_alamat, threshold):
15     hasil = process.extractOne(row, daftar_alamat, scorer=fuzz.ratio)
16     if hasil and hasil[1] >= threshold:
17         return hasil[0], hasil[1]
18     return None, None
19
20 # Mengambil Alamat dan NPWP dari file2
21 daftar_alamat2 = file2[kolom_alamat2].tolist()
22 npwp_dict = dict(zip(file2[kolom_alamat2], file2[kolom_npwp]))
23
24 # Melakukan pencocokan fuzzy untuk setiap alamat di file1
25 hasil_matching = file1[kolom_alamat1].apply(lambda x: fuzz_match(x, daftar_alamat2, 50))
26
27 # Mencetak dua kolom baru yakni alamat cocok dan skor kecocokan
28 file1["Alamat Cocok"], file1["Skor"] = zip(*hasil_matching)
29
30 # Menambahkan kolom NPWP berdasarkan hasil fuzzy matching
31 file1["NPWP"] = file1["Alamat Cocok"].map(npwp_dict)
32
33 kolom_urut = [col for col in file1.columns if col != "Skor"] + ["Skor"]
34 file1 = file1[kolom_urut]
35
36 # Hasil pencocokan
37 file1.to_csv("hasil_matching.csv", index=False)
38
39 matching_80 = file1[file1["Skor"] >= 80]
40 matching_70 = file1[(file1["Skor"] >= 70) & (file1["Skor"] < 80)]
41 matching_50 = file1[(file1["Skor"] >= 50) & (file1["Skor"] < 70)]
42
43 print(f"===== Hasil Fuzzy Matching =====")
44 print(f"Akurasi >= 80% (Total: {len(matching_80)})")
45 print(matching_80)
46 matching_80.to_csv("hasil_matching_80.csv", index=False)
47
48 print(f"Akurasi >= 70% (Total: {len(matching_70)})")
49 print(matching_70)
50 matching_70.to_csv("hasil_matching_70.csv", index=False)
51
52 print(f"Akurasi >= 50% (Total: {len(matching_50)})")
53 print(matching_50)
54 matching_50.to_csv("hasil_matching_50.csv", index=False)
55 file1
```

**Gambar 3. 3.** Program Fuzzy Matcing

**Sumber:** Dokumentasi Pribadi

Penulis melakukan proses *Fuzzy Matching* untuk mencocokkan data restoran hasil *scraping* dari Google Maps pada wilayah Kecamatan Sukolilo, yang mencakup beberapa kelurahan seperti Medokan Semampir, Semolowaru, dan Keputih, dengan memanfaatkan *tools Extensions Instant Data Scraper*. Proses pencocokan ini bertujuan untuk memastikan kesesuaian antara data alamat dan nama restoran dari kedua sumber data dengan tingkat kemiripan yang bervariasi, yaitu sebesar 80%, 70%, dan 50%. Berdasarkan hasil yang diperoleh,



menjadi lebih sistematis, sehingga dapat meminimalkan ketidaksesuaian data di lingkungan administrasi perpajakan.

### **3.3.7. Menginputkan Data Notaris untuk Pajak Pembuatan Akta Tanah (PPAT)**

Dalam administrasi perpajakan, setiap transaksi pengalihan hak atas tanah dan/atau bangunan wajib dikenakan Bea Perolehan Hak atas Tanah dan Bangunan (BPHTB) yang pembayarannya dilakukan oleh pihak penerima hak. Salah satu dokumen penting yang menjadi dasar pengenaan pajak tersebut adalah Akta Jual Beli (AJB) atau akta-akta lain yang dibuat oleh Pejabat Pembuat Akta Tanah (PPAT) atau notaris. Data dari notaris yang berkaitan dengan pembuatan akta tanah harus dicatat dan dilaporkan ke Kantor Pelayanan Pajak (KPP) sebagai bagian dari pengawasan dan pendataan objek pajak. Proses input data ini dilakukan melalui sistem administrasi pajak agar setiap transaksi pengalihan hak dapat tercatat dengan baik dan menjadi dasar dalam penerbitan surat ketetapan pajak atau validasi BPHTB. Selain itu, pendataan ini penting untuk memastikan kepatuhan wajib pajak dan menghindari potensi kehilangan penerimaan negara dari sektor pajak properti.

### **3.3.8. Input Laporan Hasil Kegiatan Pengumpulan Data ke Akses Pajak**

Dalam rangka meningkatkan kualitas basis data perpajakan, Direktorat Jenderal Pajak melakukan berbagai kegiatan pengumpulan data dari pihak ketiga maupun hasil kegiatan di lapangan. Data yang diperoleh dari pengumpulan tersebut harus dilaporkan dan diinput ke dalam aplikasi Akses Pajak, yaitu sistem internal DJP untuk mengelola, menyimpan, dan memanfaatkan data hasil pengumpulan tersebut. Laporan hasil kegiatan pengumpulan data berfungsi sebagai dokumentasi administratif atas kegiatan pengawasan yang telah dilakukan oleh pegawai KPP, baik berupa data transaksi, data objek pajak, maupun data pihak ketiga. Data tersebut kemudian diproses lebih lanjut sebagai bahan analisis kepatuhan pajak, deteksi potensi penerimaan, atau tindak

lanjut pemeriksaan. Penginputan laporan ke sistem Akses Pajak juga bertujuan untuk memperbarui database wajib pajak secara berkala dan menjadi acuan dalam pembuatan strategi pengawasan dan pemeriksaan selanjutnya. Proses ini merupakan bagian dari pelaksanaan pengawasan berbasis data yang saat ini menjadi fokus reformasi perpajakan di Indonesia

### **3.3.9. Klasifikasi Sentimen Masyarakat Terhadap Layanan**

Penulis merancang dan mengembangkan model klasifikasi sentimen menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) untuk menganalisis ulasan masyarakat terkait pelayanan di KPP Pratama Surabaya Gubeng yang diperoleh dari Google Maps dan penyebaran kuesioner. Selain itu, pengumpulan data dilakukan melalui survei kuesioner guna menghimpun pendapat serta tanggapan langsung dari masyarakat mengenai kualitas layanan perpajakan. Hasil klasifikasi sentimen ini dimanfaatkan untuk memetakan persepsi publik terhadap layanan yang diberikan, sekaligus menjadi bahan evaluasi bagi instansi. Temuan dari penelitian ini diharapkan dapat menjadi dasar dalam perumusan strategi perbaikan dan peningkatan kualitas pelayanan perpajakan yang lebih responsif dan sesuai dengan kebutuhan masyarakat di masa mendatang.

### **3.4. Korelasi Kegiatan KP dengan Mata Kuliah**

Kegiatan Kerja Praktik (KP) yang dilakukan di KPP Pratama Surabaya Gubeng memiliki keterkaitan dengan beberapa mata kuliah yang telah penulis tempuh selama mengikuti perkuliahan di program studi D-IV Sains Data Terapan. Berikut mata kuliah yang berkorelasi dengan kegiatan Kerja Praktik:

#### **1. Text Mining**

Memberikan dasar teori dan praktik dalam ekstraksi informasi dari data teks, khususnya saat melakukan analisis sentimen terhadap ulasan masyarakat tentang layanan KPP Pratama Surabaya Gubeng, serta dalam proses pembersihan dan pengolahan teks ulasan tersebut.

2. **Pemrosesan Data**  
Berperan penting dalam tahap *preprocessing data*, mulai dari pembersihan data, konversi data teks menjadi numerik, hingga sebelum digunakan dalam model klasifikasi.
3. **Machine Learning**  
Sangat relevan dalam pembuatan model klasifikasi menggunakan algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*, serta dalam melakukan evaluasi model klasifikasi yang diterapkan selama penelitian.
4. **Pemrograman**  
Menjadi dasar dalam pembuatan program, mulai dari proses *scraping data*, *preprocessing database*, hingga implementasi algoritma klasifikasi. Bahasa pemrograman Python menjadi bahasa utama yang digunakan dalam kegiatan Kerja Praktik ini.
5. **Workshop Analisis Sosial Media**  
Memberikan pengalaman praktis dalam menganalisis data sosial media atau ulasan digital, yang relevan saat melakukan pengumpulan dan analisis sentimen masyarakat terhadap layanan perpajakan melalui Google Maps.
6. **Bahasa Indonesia**  
Berperan penting dalam penyusunan laporan Kerja Praktik ini, khususnya dalam penulisan yang sistematis, penggunaan bahasa yang baik dan benar, serta penyusunan kalimat yang efektif dan sesuai kaidah penulisan ilmiah.

Halaman ini sengaja dikosongkan



## **BAB 4**

### **ANALISA DAN PEMBAHASAN**

#### **4.1. Klasifikasi Sentimen Masyarakat Terhadap Layanan KPP Pratama Surabaya Gubeng Berbasis Naive Bayes dan SVM**

Pada bab ini akan dijelaskan mengenai topik yang dibahas dalam penelitian, yaitu tentang klasifikasi sentimen masyarakat terhadap layanan yang diberikan oleh KPP Pratama Surabaya Gubeng. Penelitian ini dilakukan dengan tujuan untuk mengetahui persepsi masyarakat, khususnya para Wajib Pajak, terhadap pelayanan yang telah disediakan oleh instansi tersebut. Persepsi masyarakat sangat penting untuk dipetakan guna meningkatkan mutu layanan, mengetahui kendala di lapangan, serta memahami kebutuhan dan harapan dari pengguna layanan pajak.

Dalam penelitian ini, data dikumpulkan melalui scraping komentar masyarakat dari *platform* media sosial Google Maps dan penyebaran kuesioner digital melalui Google Forms, kemudian dilakukan proses klasifikasi sentimen menjadi kategori positif, negatif, dan netral. Metode klasifikasi yang digunakan adalah *Naive Bayes* dan *Support Vector Machine* (SVM), dua algoritma supervised learning yang sering digunakan dalam pemodelan analisis sentimen karena keandalannya dalam mengolah data teks, sehingga hasilnya diharapkan dapat menjadi bahan evaluasi dan rekomendasi perbaikan layanan publik di KPP Pratama Surabaya Gubeng.

#### **4.2. Hasil dan Pembahasan**

##### **4.2.1. Dataset**

Dataset merupakan komponen fundamental dalam penelitian ini karena menjadi landasan bagi seluruh proses analisis dan pembahasan. Untuk mendapatkan gambaran yang komprehensif mengenai Klasifikasi Sentimen Masyarakat Terhadap Layanan KPP Pratama Surabaya Gubeng, penelitian ini mengintegrasikan dua jenis sumber data yang berbeda.

Sumber pertama adalah data sekunder yang diperoleh melalui metode *web scraping* dari ulasan publik di platform

Google Maps. Data ini menyediakan informasi objektif yang mencakup rating numerik dan ulasan tekstual dari masyarakat terkait pelayanan KPP Pratama Surabaya Gubeng. Sumber kedua adalah data primer yang didapatkan dari hasil penyebaran kuesioner kepada Wajib Pajak. Data ini bertujuan untuk menangkap informasi subjektif yang lebih terstruktur, seperti tingkat kepuasan dan tanggapan kualitatif mengenai pengalaman layanan yang mereka terima. Kombinasi kedua sumber data ini memungkinkan analisis sentimen yang lebih kaya dan mendalam. Penjelasan rinci mengenai hasil akuisisi data dari masing-masing sumber akan diuraikan pada sub bab selanjutnya.

#### 4.2.1.1. Hasil Scraping Google Maps

Berdasarkan metode *web scraping* yang telah diuraikan pada sub bab 3.1.2., dilakukan pengambilan data ulasan dari platform Google Maps dengan target KPP Pratama Surabaya Gubeng. Proses *scraping* yang mencakup ulasan dari tahun 2016 hingga 2025 ini berhasil mengumpulkan total **311 data ulasan**. Adapun variabel informasi yang diekstraksi dari setiap ulasan adalah sebagai berikut:

1. **Username**: Nama pengguna yang memberikan ulasan.
2. **Date**: Tanggal ulasan dipublikasikan.
3. **Rating**: Peringkat numerik yang diberikan dalam skala 1 hingga 5 bintang.
4. **Review**: Teks ulasan atau opini yang ditulis oleh pengguna.

#### 4.2.1.2. Hasil Kuesioner Google Forms

Untuk melengkapi data sekunder, penelitian ini juga menggunakan data *primer* yang diperoleh melalui metode kuesioner, sesuai penjelasan pada sub bab 3.1.4. Kuesioner disebarikan secara daring menggunakan Google Forms pada periode **20 Januari 2025 hingga 1 Juni 2025**. Dari periode tersebut, berhasil terkumpul **154 tanggapan responden** yang dinyatakan valid dan relevan untuk penelitian. Variabel-variabel utama yang diperoleh dari kuesioner adalah sebagai berikut:

1. **Timestamp:** Waktu dan tanggal pengisian kuesioner oleh responden.
2. **Status Wajib Pajak:** Pertanyaan konfirmasi untuk memastikan responden terdaftar di KPP Pratama Surabaya Gubeng.
3. **Nama:** Identitas responden (dianonimkan dalam proses analisis).
4. **Rating:** Penilaian kuantitatif yang diberikan responden terhadap pelayanan dengan skala 1-5.
5. **Tanggapan:** Umpan balik atau masukan kualitatif dalam bentuk teks mengenai pengalaman pelayanan.

#### 4.2.2. Proses Preprocessing Data

Setelah dataset berhasil dikumpulkan dari berbagai sumber, tahap selanjutnya adalah pra-pemrosesan data (*preprocessing*). Tahap ini krusial untuk membersihkan, mentransformasi, dan menyiapkan data mentah agar menjadi format yang sesuai dan berkualitas tinggi untuk diolah pada tahap analisis sentimen. Proses pra-pemrosesan dalam penelitian ini meliputi beberapa langkah, yang akan diuraikan secara rinci pada sub bab berikut

##### 4.2.2.1. Penggabungan Data

Langkah pertama adalah menggabungkan dataset yang berasal dari Google Maps dengan data hasil kuesioner Google Forms menjadi satu. Tujuan dari tahap ini adalah untuk menciptakan satu dataset tunggal yang berisi ulasan dari kedua sumber dengan struktur yang seragam. Proses ini diawali dengan harmonisasi skema, di mana kolom-kolom dari data kuesioner disesuaikan agar selaras dengan data Google Maps. Secara spesifik, kolom 'Tanggapan' diubah namanya menjadi 'Review', kolom 'Rating' dipertahankan, dan kolom 'Timestamp' diekstrak dan diubah formatnya menjadi 'Date'. Kolom identitas responden dijadikan satu sebagai 'Username', sementara kolom lain yang tidak relevan untuk analisis sentimen diabaikan. Hasil penggabungan ini menghasilkan dataset utuh sebanyak **465 data**. Struktur final dari dataset gabungan disajikan pada Tabel 4.1.

**Tabel 4. 1.** Info Data

| Variabel | Tipe Data | Deskripsi  |
|----------|-----------|--|
| Username | Object    | Nama pengguna yang memberikan ulasan.                            |
| Date     | Date      | Tanggal ulasan dipublikasikan.                                   |
| Rating   | int64     | Peringkat numerik yang diberikan dalam skala 1 hingga 5 bintang. |
| Review   | Object    | Ulasan atau opini yang ditulis oleh pengguna.                    |

Untuk memberikan gambaran yang lebih konkret, Tabel 4.2 menyajikan 10 contoh data asli dari dataset tersebut guna menunjukkan isi dan struktur data yang akan digunakan dalam proses analisis selanjutnya.

**Tabel 4. 2.** Sampel Data Responden

|   | Username         | Date       | Rating | Review   |
|---|------------------|------------|--------|--|
| 0 | Adinda Aenun     | 30/05/2025 | 4      | Pelayanan di KPP Pratama Surabaya Gubeng sanga...  |
| 1 | cia aulia        | 30/05/2025 | 5      | Pelayanan di KPP Pratama Surabaya Gubeng sanga...  |
| 2 | Riska Saputri    | 30/05/2025 | 5      | Saya sangat mengapresiasi pelayanan yang diber...  |
| 3 | grace mega       | 30/05/2025 | 5      | Saya merasa puas dengan pelayanan di KPP Prata...  |
| 4 | ramadhan saputra | 30/05/2025 | 4      | Stafnya profesional dan sabar dalam menjawab p...  |
| 5 | Ananda Kurniawan | 30/05/2025 | 5      | Sebagai wajib pajak yang terdaftar di KPP Prat...  |
| 6 | gugun gunawan    | 30/05/2025 | 5      | Awalnya saya bingung soal e-filing, tapi petug...  |
| 7 | Very Firmansyah  | 30/05/2025 | 5      | Proses administrasi yang biasanya rumit menjad...  |
| 8 | arie ilham       | 30/05/2025 | 3      | Cukup memuaskan, hanya saja waktu tungguanya ka... |
| 9 | neng sari        | 30/05/2025 | 4      | Saya mengurus NPWP baru dan semua prosesnya be...  |

#### 4.2.2.2. *Cleaning*

Tahap *cleaning* bertujuan untuk mengatasi masalah seperti ketidakkonsistenan format, data duplikat, dan data yang tidak relevan (*noise*). Proses ini diawali dengan penggabungan dan penyeragaman data dari Google Maps dan Google Forms. Hasil pengecekan menunjukkan terdapat total 114 baris ulasan yang duplikat. Analisis terhadap ulasan duplikat tersebut disajikan pada Tabel 4.3, menunjukkan bahwa mayoritas duplikasi, yaitu sebanyak 108 kali, berasal dari ulasan yang tidak memiliki komentar dan hanya berisi simbol "-" (*hyphen*).

**Tabel 4. 3.** Ulasan Duplikat

| Review yang duplikat beserta jumlah kemunculannya: |   |                   |
|--|---|-------------------|
|  | Review  | Jumlah Kemunculan |
| 0  | -   | 108               |
| 1  | Bagus   | 3                 |
| 2  | Selamat Pagi, terima kasih untuk ulasan baikny... | 2                 |
| 3  | Pelayanan bagus                                   | 2                 |
| 4  | Tidak puas, saya datang pagi tapi selesai hamp... | 2                 |
| 5  | Kantor yang bagus                                 | 2                 |
| 6  | Sistem informasi di loket sudah oke, cuma kura... | 2                 |

Ulasan yang hanya berisi simbol "-" tidak mengandung informasi tekstual apa pun sehingga tidak dapat dianalisis sentimennya. Data ini dianggap sebagai noise yang dapat mengganggu kualitas model. Oleh karena itu, diputuskan untuk menghapus (*drop*) seluruh 108 baris data yang memiliki ulasan "-" tersebut. Setelah proses pembersihan ini, dataset menjadi lebih berkualitas dan siap untuk melalui tahap pra-pemrosesan teks selanjutnya.

#### 4.2.2.3. *Case Folding*

*Case folding* adalah tahap awal dalam pra-pemrosesan teks yang bertujuan untuk menyeragamkan karakter dalam data teks. Pada tahap ini, seluruh huruf dalam kolom 'Review' diubah menjadi huruf kecil (*lowercase*). Tujuan dari proses ini adalah untuk menghindari adanya kata yang dianggap berbeda oleh sistem hanya karena perbedaan penggunaan huruf kapital, misalnya "Bagus", "bagus", dan "BAGUS" akan dianggap sebagai satu kata yang sama. Hal ini penting untuk mengurangi dimensi data dan memastikan konsistensi pada tahap pengolahan selanjutnya.

#### 4.2.2.4. *Normalization*

Normalisasi adalah proses untuk mengubah kata-kata yang tidak baku seperti singkatan, salah ketik (*typo*), atau bahasa gaul menjadi bentuk bakunya sesuai dengan kaidah bahasa Indonesia. Proses ini krusial karena ulasan dari pengguna seringkali mengandung kata-kata non-standar.

Pada penelitian ini, normalisasi dilakukan dengan memanfaatkan sebuah kamus kata tidak baku yang telah disusun sebelumnya. Setiap kata dalam ulasan akan diperiksa dan dicocokkan dengan kamus tersebut. Jika ditemukan kata yang tidak baku, kata tersebut akan otomatis digantikan dengan bentuk bakunya.

4.2.2.5. Tokenizing

*Tokenizing* adalah proses pemecahan kalimat atau teks menjadi unit-unit yang lebih kecil yang disebut "token". Umumnya, token ini berupa kata-kata individual. Tahap ini mengubah data dari format kalimat menjadi format daftar kata, yang memungkinkan setiap kata dianalisis secara terpisah. Proses ini menjadi dasar untuk tahap selanjutnya seperti *stopword removal* dan *stemming*. Tabel 4.4 menyajikan hasil proses *case folding* hingga *tokenizing* yang siap untuk dianalisis lebih lanjut.

Tabel 4. 4. Hasil Tokenizing

|   | Review   | Normalization  | Tokenizing  |
|---|--|--|---|
| 0 | pelayanan di kpp pratama surabaya<br>gubeng sanga... | pelayanan di kpp pratama surabaya<br>gubeng sanga... | pelayanan, di, kpp, pratama,<br>surabaya, gubeng,...  |
| 1 | pelayanan di kpp pratama surabaya<br>gubeng sanga... | pelayanan di kpp pratama surabaya<br>gubeng sanga... | pelayanan, di, kpp, pratama,<br>surabaya, gubeng,...  |
| 2 | saya sangat mengapresiasi<br>pelayanan yang diber... | saya sangat mengapresiasi<br>pelayanan yang diber... | saya, sangat, mengapresiasi,<br>pelayanan, yang, ...  |
| 3 | saya merasa puas dengan<br>pelayanan di kpp prata... | saya merasa puas dengan pelayanan<br>di kpp prata... | saya, merasa, puas, dengan,<br>pelayanan, di, kpp,... |
| 4 | stafnya profesional dan sabar dalam<br>menjawab p... | stafnya profesional dan sabar dalam<br>menjawab p... | stafnya, profesional, dan, sabar,<br>dalam, menja...  |

4.2.2.6. Stopword Removal

*Stopword removal* adalah proses untuk memfilter dan menghapus kata-kata umum (*stopwords*) yang sering muncul namun tidak memiliki makna atau sentimen yang signifikan. Kata-kata seperti "dan", "yang", "di", "atau", "ini" termasuk dalam kategori *stopword*. Penghapusan kata-kata ini bertujuan untuk mengurangi *noise* dalam data dan memungkinkan model untuk fokus pada kata-kata yang benar-benar mengandung sentimen. Proses ini dilakukan dengan membandingkan setiap token dengan daftar *stopword* bahasa Indonesia. Jika sebuah

token ditemukan dalam daftar tersebut, maka token itu akan dihapus.

4.2.2.7. *Stemming*

*Stemming* adalah tahap terakhir dalam pra-pemrosesan teks pada penelitian ini. Proses ini bertujuan untuk mengubah setiap kata menjadi bentuk kata dasarnya (*root word*) dengan cara menghilangkan semua imbuhan (awalan, akhiran, sisipan, dan konfiks). Misalnya, kata "pelayanannya", "melayani", dan "dilayani" akan diubah menjadi kata dasarnya, yaitu "layan".

Tujuan utama *stemming* adalah untuk mengurangi variasi dari kata yang sama sehingga kata-kata dengan makna inti yang serupa dapat diproses sebagai satu fitur tunggal. Pada penelitian ini, proses *stemming* dilakukan menggunakan pustaka Sastrawi yang merupakan *stemmer* standar untuk Bahasa Indonesia. Setelah melalui seluruh rangkaian pra-pemrosesan ini, data teks ulasan telah bersih dan terstruktur dalam bentuk kata dasar. Dataset inilah yang siap digunakan sebagai input untuk tahap ekstraksi fitur dan pemodelan klasifikasi sentimen. Hasil proses pra-pemrosesan teks disajikan pada Tabel 4.5.

Tabel 4. 5. Hasil Stemming

|   | Stopword   | Stemming  |
|---|--|---|
| 0 | pelayanan, kpp, pratama, surabaya, gubeng, ram...  | layan, kpp, pratama, surabaya, gubeng, ramah, ... |
| 1 | kantornya, bagus, stafnya, ramah, berkunjung, ...  | kantor, bagus, staf, ramah, kunjung, antri, sa... |
| 2 | netral, berjalan, sesuai, prosedur, efisien        | netral, jalan, sesuai, prosedur, efisien          |
| 3 | pelayanannya, cepat, berbelit, belit, tempatny...  | layan, cepat, belit, belit, tempat, bagus, sih... |
| 4 | ruang, tungguanya, nyaman, parkir, motor, mobil... | ruang, tunggu, nyaman, parkir, motor, mobil, s... |

4.2.3. Pemetaan Topik Ulasan

Setelah teks ulasan melalui seluruh tahap pra-pemrosesan, langkah selanjutnya adalah melakukan pemetaan untuk mengkategorikan setiap ulasan berdasarkan topik yang dibahas. Proses ini bertujuan untuk mengidentifikasi aspek apa yang paling sering menjadi fokus dalam ulasan masyarakat. Pemetaan dilakukan menggunakan pendekatan berbasis aturan (*rule-based*) dengan mendeteksi kata kunci (*keywords*) di dalam setiap ulasan

yang telah di-stemming. Terdapat dua kategori utama yang didefinisikan:

1. Pelayanan: Jika ulasan mengandung kata kunci yang berhubungan dengan layanan (contoh: 'layan', 'cepat', 'ramah', 'bantu', 'proses').
2. Tempat: Jika ulasan mengandung kata kunci yang terkait dengan fasilitas fisik (contoh: 'ruang', 'parkir', 'bersih', 'nyaman').
3. Jika sebuah ulasan tidak mengandung kata kunci dari kedua kategori tersebut, maka ulasan akan diklasifikasikan sebagai 'Overall', yang diasumsikan memberikan tanggapan umum.

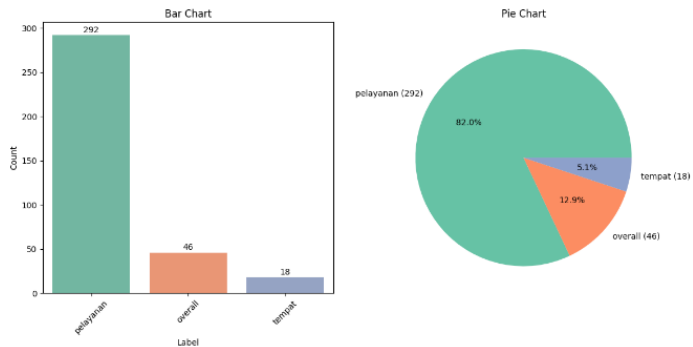
Untuk memahami distribusi topik yang dibahas dalam ulasan masyarakat, dilakukan pemetaan jumlah ulasan berdasarkan label topik. Berdasarkan proses pemetaan ini, diperoleh distribusi topik ulasan disajikan pada Tabel 4.6.

**Tabel 4. 6.** Pemetaan Topik

|   | <b>Label</b>     | <b>Jumlah</b> |
|---|------------------|---------------|
| 0 | <b>pelayanan</b> | 292           |
| 1 | <b>overall</b>   | 46            |
| 2 | <b>tempat</b>    | 18            |

Hasil pada Tabel 4.6 menunjukkan bahwa *pelayanan* merupakan topik yang paling dominan, dengan jumlah sebanyak 292 ulasan. Hal ini mengindikasikan bahwa kualitas layanan menjadi perhatian utama dari para Wajib Pajak dalam memberikan tanggapan atau penilaian. Visualisasi persebaran topik ini ditampilkan pada Gambar 4.1 dalam bentuk *bar chart* dan *pie chart*, yang semakin memperjelas dominasi topik pelayanan dibandingkan dengan topik-topik lain.



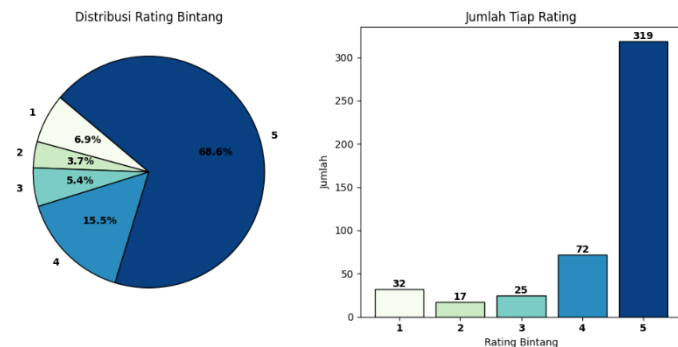


**Gambar 4. 1.** Chart Pemetaan Topik

**Sumber:** Dokumentasi Pribadi

#### 4.2.4. Eksplorasi Data

Tahap awal dalam analisis adalah melakukan eksplorasi data untuk memahami karakteristik dan distribusi dari dataset ulasan yang telah dikumpulkan. Visualisasi dari distribusi ini disajikan pada Gambar 4.2, yang terdiri dari diagram lingkaran untuk menunjukkan proporsi dan diagram batang untuk menunjukkan jumlah absolut dari setiap kategori rating bintang.



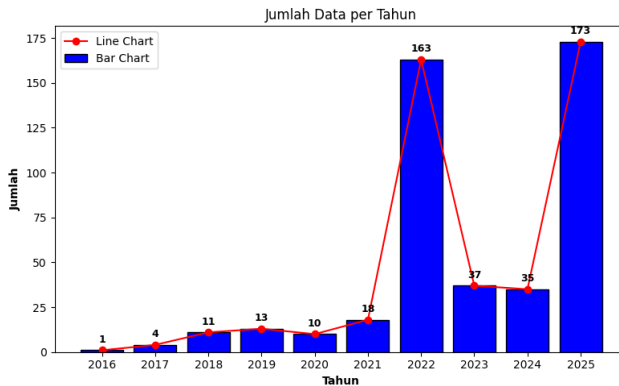
**Gambar 4. 2.** Visualisasi Distribusi Rating Bintang

**Sumber:** Dokumentasi Pribadi

Berdasarkan kedua grafik tersebut, terlihat jelas bahwa dataset memiliki distribusi yang sangat tidak seimbang dan cenderung sangat positif. Dari total 465 ulasan yang dianalisis,

rating bintang 5 mendominasi secara signifikan dengan jumlah 319 ulasan, atau mencakup 68,6% dari keseluruhan data. Rating bintang 4 menempati posisi kedua dengan 72 ulasan (15,5%), yang semakin memperkuat dominasi sentimen positif dalam dataset ini. Hal yang menarik untuk dicatat adalah pola distribusi pada rating rendah. Jumlah rating bintang 1 (32 ulasan atau 6,9%) ternyata lebih tinggi dibandingkan dengan rating bintang 2 (17 ulasan atau 3,7%) dan rating bintang 3 (25 ulasan atau 5,4%). Pola ini menunjukkan bahwa pengguna lebih cenderung memberikan ulasan pada saat mereka merasa sangat puas (bintang 5) atau sangat tidak puas (bintang 1), dibandingkan dengan pengalaman yang moderat atau netral.

Untuk memahami distribusi data dari sisi waktu, dilakukan analisis terhadap jumlah ulasan yang terkumpul setiap tahunnya dari periode 2016 hingga 2025. Gambar 4.3 memvisualisasikan tren dan volume data tahunan ini melalui kombinasi diagram batang dan garis untuk menyoroti fluktuasi yang terjadi.



**Gambar 4. 3.** Jumlah Data Ulasan per Tahun

**Sumber:** Dokumentasi Pribadi

Dari grafik tersebut, dapat diamati bahwa volume data ulasan tidak konsisten setiap tahunnya. Pada periode awal (2016-2021), jumlah ulasan yang masuk relatif rendah, dengan angka tertinggi hanya mencapai 18 ulasan pada tahun 2021. Hal ini

menunjukkan tingkat interaksi atau popularitas yang masih terbatas pada tahun-tahun tersebut. Namun, terjadi lonjakan yang sangat drastis pada tahun 2022, di mana jumlah ulasan meroket hingga 163 ulasan. Setelah puncak tersebut, volume data mengalami normalisasi di angka yang lebih tinggi dari periode awal, yaitu 37 ulasan pada tahun 2023 dan 35 ulasan pada tahun 2024.

Tren lonjakan signifikan kembali terjadi dan bahkan mencapai puncaknya pada tahun 2025, dengan total 173 ulasan terkumpul. Konsentrasi data yang sangat tinggi pada tahun 2022 dan 2025 ini merupakan temuan penting. Hal ini mengindikasikan bahwa analisis sentimen yang akan dilakukan kemungkinan besar akan sangat dipengaruhi oleh persepsi atau peristiwa yang terjadi pada kedua tahun tersebut. Adanya fluktuasi ekstrem ini menunjukkan bahwa sentimen publik bisa jadi tidak statis, melainkan dipengaruhi oleh faktor-faktor eksternal tertentu seperti perubahan layanan, promosi, atau kejadian viral pada tahun-tahun puncak tersebut.

#### **4.2.5. Pemodelan dan Ekperimen**

Setelah melalui eksplorasi data, langkah selanjutnya adalah implementasi model klasifikasi untuk menentukan sentimen dari setiap ulasan. Tahap ini mencakup pelabelan sentimen, penyeimbangan dataset, pembangunan model dengan algoritma *Naive Bayes* dan *Support Vector Machine*, serta evaluasi performa masing-masing model.

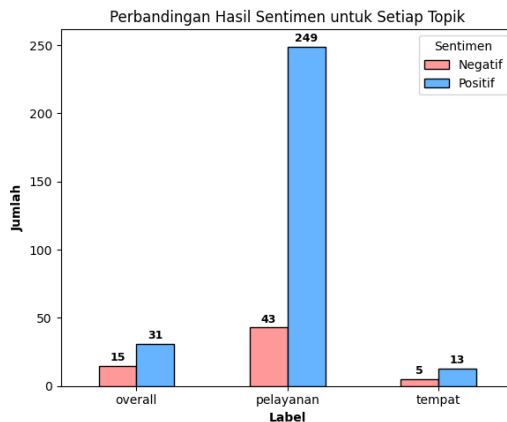
##### **4.2.5.1. Pelabelan Sentimen**

Pelabelan sentimen awal dilakukan dengan menggunakan pendekatan hibrida yang menggabungkan informasi rating numerik dengan analisis polaritas teks menggunakan pustaka *TextBlob*. Aturan pelabelan yang diterapkan adalah sebagai berikut:

1. Jika Rating bernilai 4 atau 5, ulasan secara otomatis diberi label 'Positif'.

2. Jika Rating bernilai 1 atau 2, ulasan secara otomatis diberi label 'Negatif'.
3. Jika Rating bernilai 3 (dianggap netral atau ambigu), sentimen ditentukan berdasarkan polaritas teks dari Review yang dianalisis oleh TextBlob.
4. Jika polaritas  $> 0$ , labelnya adalah 'Positif'; jika tidak, labelnya adalah 'Negatif'.

Proses pelabelan sentimen terhadap data ulasan menghasilkan distribusi awal yang tidak merata antara dua kelas utama, yaitu positif dan negatif. Berdasarkan hasil pelabelan, terdapat sebanyak **293 komentar positif** dan **63 komentar negatif**. Ketimpangan ini mencerminkan adanya fenomena *class imbalance* yang cukup signifikan, di mana jumlah data pada kelas positif jauh lebih dominan dibandingkan kelas negatif. Visualisasi dari distribusi ini dapat dilihat pada **Gambar 4.4**, yang disajikan dalam bentuk *bar chart* untuk memberikan gambaran yang lebih jelas mengenai ketidakseimbangan kelas dalam dataset.



**Gambar 4. 4.** Chart Distribusi Sentimen

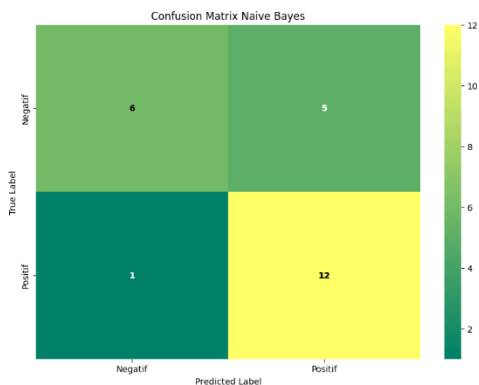
**Sumber:** Dokumentasi Pribadi

#### 4.2.5.2. Penyeimbangan Dataset

Ketidakseimbangan kelas dapat menyebabkan model menjadi bias dan cenderung memprediksi kelas mayoritas. Untuk mengatasi masalah ini, dilakukan teknik penyeimbangan dataset dengan metode *undersampling*. Metode ini bekerja dengan cara mengurangi jumlah sampel dari kelas mayoritas 'Positif' agar setara dengan jumlah sampel dari kelas minoritas 'Negatif'. Dalam penelitian ini, dilakukan pengambilan sampel acak sebanyak 60 data dari kelas 'Positif' dan 60 data dari kelas 'Negatif'. Hasilnya adalah dataset baru yang seimbang (*balanced*) dengan total 120 data, yang siap digunakan untuk melatih dan menguji model.

#### 4.2.5.3. Klasifikasi Menggunakan Model Naive Bayes (NB)

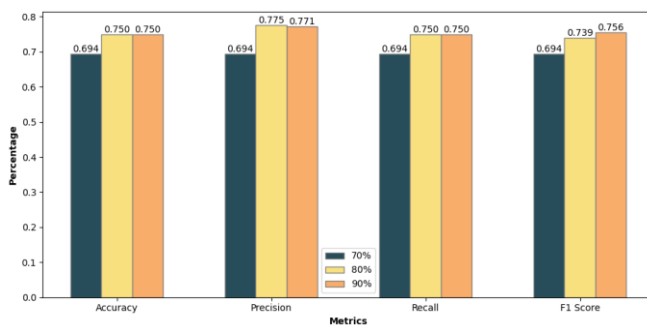
Model klasifikasi pertama yang diimplementasikan dalam proyek ini adalah **Naive Bayes**. Dataset yang telah mengalami penyeimbangan kelas terlebih dahulu dibagi menjadi data latih dan data uji dengan proporsi **80:20**, yaitu sebanyak **96 data latih** dan **24 data uji**. Proses pelatihan dan pengujian model menghasilkan *confusion matrix* seperti yang ditampilkan pada Gambar 4.5.



**Gambar 4. 5.** Confussion Matrix NB

**Sumber:** Dokumentasi Pribadi

Berdasarkan hasil confusion matrix, diketahui bahwa model Naive Bayes mampu mengklasifikasikan data dengan tingkat ketepatan yang bervariasi antar kelas. Untuk kelas Negatif, model menunjukkan tingkat akurasi sebesar 63.64%, sedangkan untuk kelas Positif, akurasi yang dicapai adalah 61.54%. Hal ini menunjukkan bahwa meskipun model memiliki kemampuan dasar dalam membedakan kedua kelas, masih terdapat ruang untuk perbaikan terutama dalam hal keseimbangan prediksi. Untuk menguji konsistensi dan stabilitas performa model, dilakukan pengujian tambahan dengan menggunakan beberapa skenario proporsi data pelatihan, yaitu 70%, 80%, dan 90%. Setiap skenario diukur menggunakan metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-score*. Untuk menunjukkan performa terbaik, pengujian dengan **proporsi 90:10** dipilih sebagai acuan visualisasi. Rangkuman hasil evaluasi performa model Naive Bayes dari ketiga skenario tersebut disajikan pada Gambar 4.6.



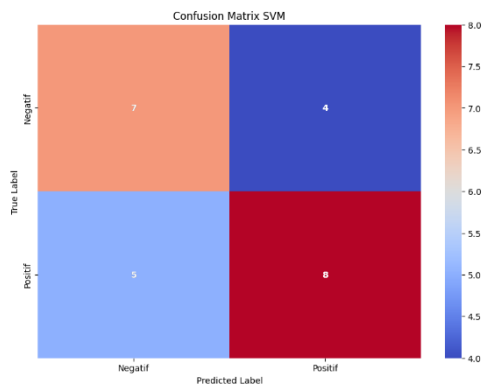
**Gambar 4. 6.** Evaluasi Performansi Model Naive Bayes

**Sumber:** Dokumentasi Pribadi

#### 4.2.5.4. Klasifikasi Menggunakan Support Vector Machine (SVM)

Model kedua yang diimplementasikan adalah *Support Vector Machine* (SVM) dengan alur pemrosesan yang sama seperti pada model *Naive Bayes*. Dataset yang telah diseimbangkan dibagi menjadi **96 data latih** dan **24 data uji**

(proporsi 80:20). Visualisasi hasil prediksi ditunjukkan dalam bentuk *confusion matrix* pada Gambar 4.7, yang memperlihatkan distribusi prediksi model terhadap data uji. Diperoleh hasil bahwa akurasi untuk kelas Negatif adalah 63.64%, sedangkan akurasi untuk kelas Positif mencapai 61.54%. Nilai ini mencerminkan performa model yang cukup seimbang, meskipun masih memiliki potensi untuk peningkatan lebih lanjut.

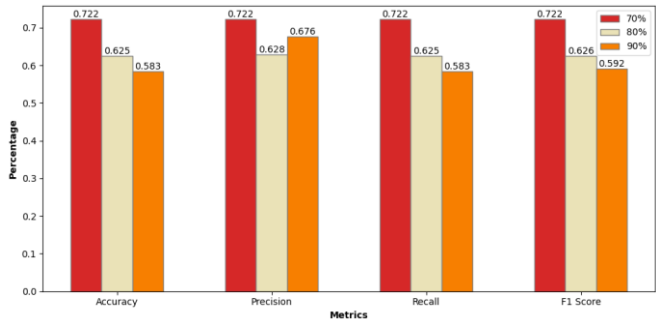


**Gambar 4. 7.** Confusion Matrix SVM

**Sumber:** Dokumentasi Pribadi

Selanjutnya, dilakukan evaluasi performa model dengan menggunakan tiga skenario proporsi data pelatihan, yaitu 70%, 80%, dan 90%, untuk mengukur pengaruh jumlah data latih terhadap kinerja model. Berdasarkan hasil pengujian, proporsi 70:30 memberikan hasil terbaik. dengan akurasi mencapai 72,22%. Terjadi tren penurunan akurasi seiring dengan penambahan jumlah data latih, di mana pada rasio 80:20 akurasi turun menjadi 62,50% dan terus menurun hingga 58,33% pada rasio 90:10. Fenomena ini dapat mengindikasikan adanya potensi *overfitting*, di mana model menjadi terlalu "hafal" dengan karakteristik data latih yang lebih besar sehingga kemampuannya untuk generalisasi pada

data uji yang baru menjadi berkurang. Detail metrik performa untuk masing-masing skenario ditampilkan pada Gambar 4.8.



**Gambar 4. 8.** Evaluasi Performansi Model SVM  
**Sumber:** Dokumentasi Pribadi

**4.2.6. Perbandingan Kinerja Model Klasifikasi**

Setelah hasil pemodelan eksperimen kedua model klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), dievaluasi untuk membandingkan kinerjanya dalam menentukan sentimen ulasan. Evaluasi dilakukan menggunakan metrik standar industri yang mencakup akurasi, presisi, *recall*, dan *F1-score*. Selain itu, analisis juga diperdalam untuk melihat kemampuan model dalam memprediksi sentimen pada topik spesifik yang berbeda. Perbandingan performa secara keseluruhan antara model *Naïve Bayes* dan SVM disajikan pada Tabel 4.7.

**Tabel 4. 7.** Perbandingan Metrik Kinerja

|          | Naïve Bayes | SVM    |
|----------|-------------|--------|
| Evaluasi |             |        |
| Akurasi  | 0.7500      | 0.7222 |
| Presisi  | 0.7714      | 0.7222 |
| Recall   | 0.7500      | 0.7222 |
| F1-score | 0.7556      | 0.7222 |



Berdasarkan hasil pada tabel di atas, terlihat bahwa model *Naïve Bayes* menunjukkan keunggulan tipis dibandingkan SVM pada seluruh metrik evaluasi. *Naïve Bayes* berhasil mencapai akurasi sebesar 75% dengan presisi 77,14%, yang mengindikasikan bahwa dari semua prediksi positif yang dibuat, sekitar 77% di antaranya benar. Performa SVM sedikit di bawahnya dengan angka yang konsisten di 72,22% untuk semua metrik, menunjukkan kinerja yang stabil namun belum seoptimal *Naïve Bayes* pada dataset ini.

Untuk memahami lebih dalam di mana keunggulan dan kelemahan masing-masing model, dilakukan analisis akurasi prediksi sentimen yang dipecah berdasarkan topik utama yang muncul dari ulasan. Hasilnya disahajikan pada Tabel 4.8 dan Tabel 4.9.

**Tabel 4. 8.** Akurasi Sentimen NB Per Topik

|   | Label     | NB Negatif | NB Positif |
|---|-----------|------------|------------|
| 0 | Pelayanan | 57.14%     | 90.91%     |
| 1 | Tempat    | 100%       | 0.00%      |
| 2 | Overall   | 100%       | 100%       |

**Tabel 4. 9.** Akurasi Sentimen SVM Per Topik

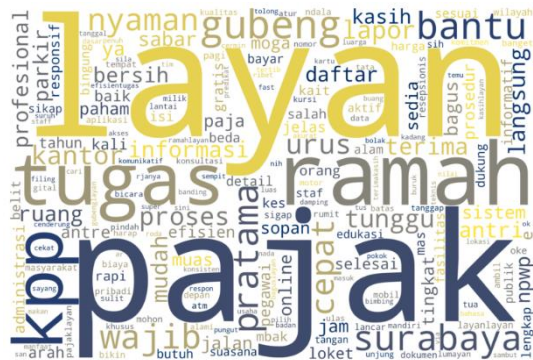
|   | Label     | SVM Negatif | SVM Positif |
|---|-----------|-------------|-------------|
| 0 | Pelayanan | 100%        | 90.91%      |
| 1 | Tempat    | 100%        | 0.00%       |
| 2 | Overall   | 100%        | 100%        |

Analisis per topik ini mengungkap temuan yang sangat signifikan. Pada topik ‘Pelayanan’, kedua model sangat baik dalam mengenali sentimen positif (90,91%), namun SVM secara superior mampu mengidentifikasi sentimen negatif (100%) dibandingkan *Naïve Bayes* yang kesulitan (57,14%).

Temuan paling krusial terlihat pada topik ‘Tempat’. Kedua model menunjukkan kegagalan dalam mengenali sentimen positif, meskipun keduanya sempurna dalam mengenali sentimen negatif. Fenomena ini kemungkinan besar disebabkan oleh ketidakseimbangan data yang ekstrem pada topik ‘Tempat’, di mana tidak ada cukup contoh ulasan positif mengenai tempat dalam data latih, sehingga model tidak mampu mempelajari polanya.

#### 4.2.7. Visualisasi Frekuensi Kata dengan Word Cloud

Visualisasi digunakan untuk memahami fokus utama dari opini yang diberikan oleh pengguna. Untuk visualisasi awal yang intuitif, digunakan metode *Word Cloud* seperti yang tersaji pada Gambar 4.9.

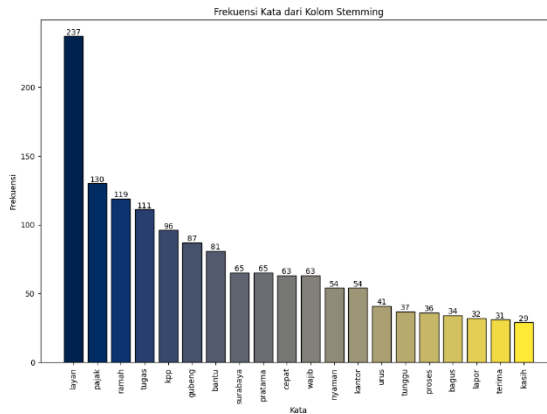


**Gambar 4. 9.** Wordcloud

**Sumber:** Dokumentasi Pribadi

Dalam visualisasi ini, ukuran sebuah kata merepresentasikan frekuensi kemunculannya dalam dataset. Terlihat jelas bahwa kata-kata seperti ‘pajak’, ‘layanan’, ‘ramah’, dan ‘cepat’ merupakan istilah yang paling dominan dan menonjol, memberikan gambaran umum mengenai topik utama yang dibicarakan. Untuk mendapatkan analisis yang lebih presisi dan terukur, dilakukan pula penghitungan frekuensi setiap kata. Hasilnya divisualisasikan dalam bentuk diagram batang pada Gambar 4.10, yang

mengurutkan kata-kata dari yang paling sering hingga yang jarang muncul.



**Gambar 4. 10.** Frekuensi Kata  
**Sumber:** Dokumentasi Pribadi

Grafik frekuensi ini mengonfirmasi temuan dari *Word Cloud* dan memberikan data kuantitatif yang lebih detail. Kata ‘layan’ (bentuk dasar dari ‘pelayanan’) menempati urutan pertama dengan frekuensi kemunculan sebanyak 237 kali. Ini menunjukkan bahwa aspek pelayanan adalah topik yang paling sentral dalam ulasan. Kata ini diikuti oleh ‘pajak’ (130 kali), ‘ramah’ (119 kali), dan ‘tugas’ (111 kali), di mana kata ‘tugas’ kemungkinan besar adalah bentuk dasar dari kata ‘petugas’, yang merujuk pada staf atau pegawai.

Secara keseluruhan, kedua visualisasi ini secara konsisten menyoroti beberapa tema utama: kualitas layanan (ditandai oleh kata layan, ramah, cepat, bantu, proses, antri), subjek utama (pajak), dan konteks spesifik yang merujuk pada personel (tugas/petugas) serta lokasi (kpp, gubeng, surabaya, pratama). Kemunculan dominan kata-kata bernuansa positif seperti ‘ramah’, ‘bantu’, ‘cepat’, ‘baik’, ‘nyaman’, dan ‘terima kasih’ memberikan indikasi awal bahwa dataset ini memiliki kecenderungan sentimen positif. Eksplorasi kata kunci ini menjadi dasar penting yang

mengonfirmasi relevansi dataset dan memberikan pemahaman awal mengenai topik-topik yang akan dianalisis lebih dalam pada tahap pemodelan klasifikasi sentimen.

#### 4.2.7.1. Analisis Frekuensi Kata pada Topik ‘Pelayananan’

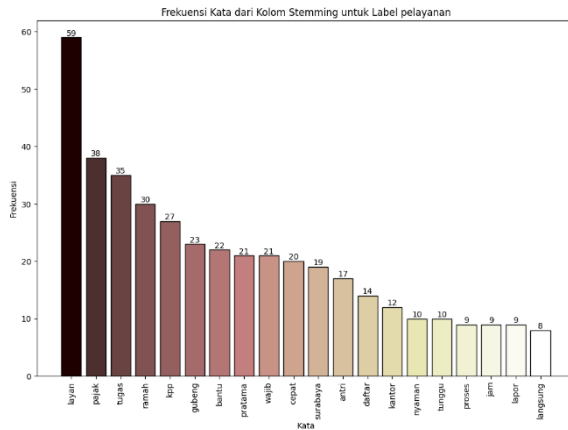
Untuk memahami konteks ulasan yang telah dipetakan ke dalam topik **‘pelayanan’**, dilakukan analisis frekuensi kata secara lebih mendalam. Visualisasi kata kunci disajikan dalam bentuk *Word Cloud* pada gambar 4.11.



**Gambar 4. 11. Wordcloud Pelayanan**

**Sumber:** Dokumentasi Pribadi

Untuk memberikan gambaran intuitif, yang kemudian diperkuat dengan diagram batang frekuensi kata untuk analisis kuantitatif yang lebih presisi, seperti yang ditunjukkan pada Gambar 4.11. Hasil analisis menunjukkan bahwa kata ‘layan’ (muncul 59 kali) dan ‘pajak’ (38 kali) tetap menjadi istilah yang paling dominan. Hal yang signifikan adalah munculnya kata-kata kunci lain yang secara spesifik menggambarkan interaksi pelayanan, seperti ‘tugas’ (merujuk pada petugas) dengan frekuensi 35 kali, diikuti oleh ‘ramah’ (30 kali), ‘bantu’ (22 kali), dan ‘cepat’ (20 kali).



**Gambar 4. 12.** Frekuensi Kata Topik Pelayanan

**Sumber:** Dokumentasi Pribadi

Temuan ini mengonfirmasi bahwa diskusi seputar topik ‘pelayanan’ sangat erat kaitannya dengan dua aspek utama: perilaku staf (diwakili kata tugas, ramah, bantu) dan efisiensi proses (diwakili kata cepat, antri, tunggu). Analisis mendalam ini memberikan pemahaman kontekstual yang kaya mengenai dimensi spesifik dari pelayanan yang dinilai oleh pengguna, yang menjadi dasar kuat untuk tahap analisis sentimen per topik selanjutnya.

#### 4.2.7.2. Analisis Frekuensi Kata pada Topik ‘Tempat’

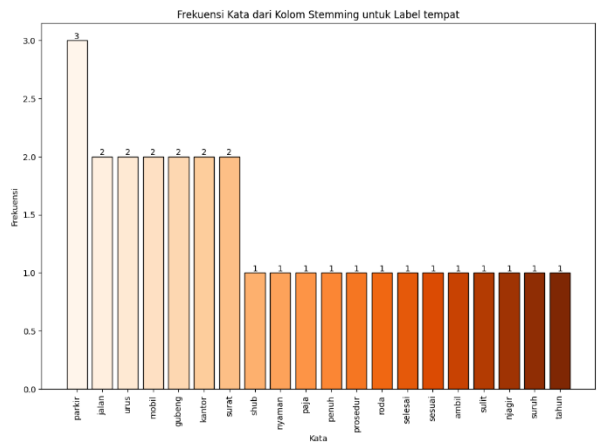
Visualisasi frekuensi kata pada label “tempat” divisualisasikan dalam bentuk *WordCloud* yang ditampilkan pada Gambar 4.13. *WordCloud* merupakan representasi visual dari teks, di mana ukuran setiap kata mencerminkan frekuensi kemunculannya dalam kumpulan data. Semakin besar ukuran suatu kata, semakin sering kata tersebut muncul dalam data.

*WordCloud* memudahkan pembaca untuk secara cepat mengidentifikasi kata-kata yang paling dominan dalam kumpulan data. Beberapa kata yang tampak dominan antara lain "parkir", "jalan", "urus", dan "mobil", yang paling menonjol, menggambarkan bahwa ulasan masyarakat yang

dikategorikan dalam label *tempat* sering kali berkaitan dengan kondisi parkiran, akses jalan, serta pelayanan administratif di kantor.



**Gambar 4. 13.** WordCloud Tempat  
**Sumber:** Dokumentasi Pribadi



**Gambar 4. 14.** Frekuensi Kata Topik Tempat  
**Sumber:** Dokumentasi Pribadi

Gambar 4.14 menyajikan diagram batang dari frekuensi kata-kata hasil *stemming* yang muncul pada label “tempat”. Diagram ini memberikan informasi kuantitatif yang lebih rinci

dibandingkan *WordCloud*, dengan menampilkan jumlah kemunculan masing-masing kata secara eksplisit.

Dari visualisasi ini dapat diketahui bahwa kata “parkir” memiliki frekuensi tertinggi yaitu sebanyak 3 kali, diikuti oleh beberapa kata lain seperti “jalan”, “urus”, “mobil”, “gubeng”, dan “kantor” yang masing-masing muncul sebanyak 2 kali. Sisanya seperti “surat”, “nyaman”, dan lain-lain muncul satu kali. Diagram ini mendukung temuan dari *WordCloud* dan menunjukkan secara detail bahwa permasalahan atau perhatian publik pada aspek tempat berkaitan dengan fasilitas kendaraan, alur jalan, dan keberadaan kantor layanan.

#### 4.2.5.1. Analisis Frekuensi Kata pada Topik ‘Overall’

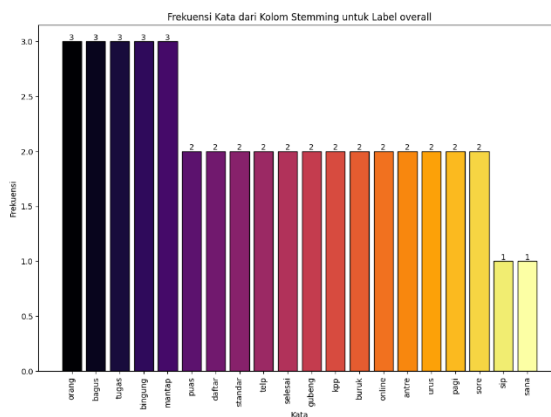
Visualisasi frekuensi kata pada label “overall” disajikan dalam dua bentuk, yaitu *WordCloud* dan diagram batang, sebagaimana ditampilkan pada Gambar 4.15. Pada diagram batang, tampak bahwa kata “orang”, “bagus”, “tugas”, “bingung”, dan “mantap” memiliki frekuensi tertinggi, masing-masing muncul sebanyak 3 kali. Disusul oleh beberapa kata lain yang muncul dua kali seperti “puas”, “daftar”, “standar”, “telp”, “selesai”, dan “buruk”. Kata-kata ini mencerminkan opini atau penilaian umum masyarakat secara keseluruhan terhadap pelayanan yang diterima.



Gambar 4. 15. WordCloud Overall

**Sumber:** Dokumentasi Pribadi

Sementara itu, *WordCloud* disajikan pada Gambar 4.16. memberikan representasi visual yang memperkuat temuan dari diagram batang. Kata-kata yang paling sering muncul ditampilkan dengan ukuran huruf yang lebih besar, sehingga memudahkan dalam mengidentifikasi tema sentimen utama secara intuitif. Misalnya, kata "bagus", "puas", dan "mantap" menunjukkan sentimen positif, sedangkan "bingung" dan "buruk" menandakan adanya beberapa pengalaman yang kurang menyenangkan.



**Gambar 4. 16.** Frekuensi Kata Topik Overall

**Sumber:** Dokumentasi Pribadi

Kombinasi kedua visualisasi ini memberikan gambaran menyeluruh tentang persepsi publik dalam label *overall*, baik dari sisi kuantitatif maupun visual-interpretatif, yang bermanfaat untuk memahami kecenderungan sentimen masyarakat terhadap pelayanan secara umum.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil pelaksanaan Kerja Praktik di KPP Pratama Surabaya Gubeng yang telah dilakukan pada 13 Januari – 9 Mei 2025, dapat disimpulkan bahwa kegiatan pengumpulan dan pengolahan data sentimen masyarakat melalui platform Google Maps dan kuesioner dari Google Forms berhasil dilakukan dengan baik. Data sentimen yang diperoleh kemudian diproses melalui tahap preprocessing dan balancing hingga menghasilkan data yang siap digunakan untuk analisis klasifikasi. Dua algoritma machine learning, yaitu **Naive Bayes** dan **Support Vector Machine (SVM)**, telah berhasil diterapkan untuk membangun model klasifikasi sentimen terhadap layanan KPP Pratama Surabaya Gubeng.

Hasil evaluasi menunjukkan bahwa kedua algoritma mampu mengklasifikasikan sentimen ke dalam kategori **positif** dan **negatif**, meskipun terdapat perbedaan performa di antara keduanya. Dari model pengujian, didapatkan bahwa **algoritma Naive Bayes memberikan performa yang lebih baik secara keseluruhan** untuk analisis sentimen pada dataset ini. Naive Bayes menunjukkan hasil yang lebih stabil pada berbagai porsi data training, khususnya pada porsi 80% dan 90%, dengan nilai **akurasi dan F1-score yang konsisten lebih tinggi dibandingkan SVM**. Sementara itu, algoritma SVM hanya unggul pada porsi training 70%, namun mengalami penurunan performa yang signifikan saat porsi data training ditingkatkan. Hal ini menunjukkan bahwa untuk karakteristik data ulasan layanan di KPP Pratama Surabaya Gubeng yang relatif sederhana dan bersifat text-based, metode probabilistik seperti Naive Bayes lebih efektif dalam mengklasifikasikan sentimen masyarakat dibandingkan SVM.

Selain analisis sentimen, penulis juga berkontribusi dalam berbagai tugas administrasi, pelayanan, pengolahan, dan analisis data di lingkungan KPP Pratama Surabaya Gubeng, mulai dari input data notaris, blast WhatsApp reminder, pembuatan database PPAT, hingga analisis risiko wajib pajak. Dengan demikian, Kerja Praktik ini

memberikan pengalaman langsung dalam penerapan data science di sektor pelayanan publik sekaligus menjadi sarana untuk mendalami penerapan metode analisis data berbasis teks di lingkungan instansi pemerintah.

## **5.2. Saran**

Berdasarkan hasil pelaksanaan kerja praktik dan *project* yang telah dilakukan, penulis memperoleh berbagai pengalaman serta wawasan baru. Namun demikian, masih terdapat sejumlah hal yang perlu diperhatikan dan ditingkatkan, baik oleh KPP Pratama Surabaya Gubeng, pihak PENS, maupun penulis sendiri. Oleh sebab itu, saran-saran berikut disusun sebagai bentuk evaluasi sekaligus rekomendasi guna mendukung peningkatan kualitas pelayanan, proses pembelajaran, dan pengembangan keilmuan di masa yang akan datang. Adapun saran yang dapat disampaikan ditujukan kepada beberapa pihak terkait sebagai berikut:

### **5.2.1. Kepada KPP Pratama Surabaya Gubeng**

Diharapkan KPP Pratama Surabaya Gubeng dapat memanfaatkan hasil analisis sentimen ini sebagai bahan evaluasi rutin untuk meningkatkan kualitas layanan publik, khususnya pelayanan administrasi perpajakan. Selain itu, pengembangan sistem monitoring opini masyarakat berbasis machine learning secara berkala di berbagai kanal media sosial dan platform ulasan publik dapat dijadikan strategi baru dalam mengelola persepsi masyarakat secara real-time.

### **5.2.2. Kepada Politeknik Elektronika Negeri Surabaya**

Agar dapat terus mendorong mahasiswa dalam pelaksanaan Kerja Praktik yang berbasis project riil di instansi pemerintahan khususnya di KPP Pratama Surabaya Gubeng atau industri, serta memperkuat mata kuliah berbasis data science seperti Text Mining, Pemrosesan Data, AI, Machine Learning, dan Analisis Media Sosial dengan penerapan proyek lapangan. Selain itu, materi penyusunan laporan berbasis hasil riset juga perlu

diintegrasikan melalui mata kuliah Bahasa Indonesia di semester awal agar mahasiswa lebih siap menghadapi laporan Kerja Praktik.

### **5.2.3. Kepada Penulis**

Penulis diharapkan dapat mengembangkan hasil penelitian ini lebih lanjut, khususnya dengan menerapkan algoritma lain seperti Random Forest, XGBoost, atau LSTM untuk klasifikasi sentimen yang lebih akurat dan adaptif terhadap data teks tidak terstruktur. Selain itu, pengalaman selama kerja praktik menjadi bekal penting bagi penulis dalam menerapkan ilmu di dunia kerja, khususnya pada bidang data science dan pelayanan publik berbasis digital.

Halaman ini sengaja dikosongkan

## DAFTAR PUSTAKA

- [1] Gupta, V., & Lehal, G. S. (2019). *A Survey of Text Mining Techniques and Applications*. *IEEE Access*, 7, 21687-21712.
- [2] Latif, S., Usama, M., Qadir, J., & Castro, I. (2021). *Leveraging Data Science and Machine Learning in Smart City Applications: A Comprehensive Survey*. *IEEE Access*, 9, 47922-47962
- [3] Bhatti, H. S., Raza, B., Khalid, S., & Iqbal, J. (2021). *A Comparative Analysis of Machine Learning Algorithms for Sentiment Analysis Using Python*. *IEEE Access*, 9, 123612-123625.
- [4] Chen, S. (2025). *Sentiment Analysis Techniques for Deep Learning Classification and Comparison*. *Theoretical and Natural Science*, 86, 74–80
- [5] Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). *Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research*. *arXiv preprint arXiv:2005.00357*.
- [6] Wikipedia contributors. (2024). *Google Maps*. In Wikipedia, The Free Encyclopedia. Retrieved from [https://en.wikipedia.org/wiki/Google\\_Maps](https://en.wikipedia.org/wiki/Google_Maps)
- [7] Derrow-Pinion, A., She, J., Wong, D., Lange, O., Hester, T., Perez, L., ... & Veličković, P. (2021). *ETA Prediction with Graph Neural Networks in Google Maps*. *arXiv preprint arXiv:2108.11482*.
- [8] ubarok, A., Aprilia, N. T., & Susanti, S. (2020). *Analisis Kepuasan Pengguna Layanan Google-Forms Sebagai Media Survey Online Menggunakan DeLone & McLean*. *Jurnal Informatika*, 7(2).
- [9] Wikipedia contributors. (2024). *Web scraping*. In Wikipedia, The Free Encyclopedia. Retrieved from [https://en.wikipedia.org/wiki/Web\\_scraping](https://en.wikipedia.org/wiki/Web_scraping)
- [10] Brenning, A., & Henn, S. (2023). *Web scraping: a promising tool for geographic data acquisition*. *arXiv preprint arXiv:2305.19893*.
- [11] Alaydaa, M. S. M., Li, J., & Jenkins, K. (2023). *Aspect based sentimental analysis for travellers' reviews*. *arXiv preprint arXiv:2308.02548*.
- [12] Aufar, A. F., Rosid, M. A., Eviyanti, A., & Astutik, I. R. I. (2023). *Mengoptimalkan Preprocessing Teks untuk Analisis Sentimen yang Akurat pada Ulasan E-Wallet*. *Journal of Information and Computer Technology Education (JICTE)*, 7(2).

- [13] Arief, M., & Deris, M. B. M. (2021). *Text Preprocessing Impact for Sentiment Classification in Product Review*. 2021 Sixth International Conference on Informatics and Computing (ICIC), 2021, pp. 1–6.
- [14] Wikipedia contributors. (2024). *Naive Bayes classifier*. In Wikipedia, The Free Encyclopedia. Retrieved from [https://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](https://en.wikipedia.org/wiki/Naive_Bayes_classifier)
- [15] Powers, D. M. W. (2020). *Evaluation: From Precision, Recall and F-Score to ROC, Informedness, Markedness & Correlation*. arXiv preprint arXiv:2010.16061.
- [16] S. Tripathy, A. Samal, D. P. Dogra, "Sentiment analysis using rule-based classification and Min-Max normalization for social media text," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 2, pp. 315-326
- [17] Behnam Yousefimehr et al., "Data Balancing Strategies: A Survey of Resampling and Augmentation Methods," *arXiv preprint arXiv:2505.13518*, 2025.

## LAMPIRAN

### Lampiran 1. Biodata Penulis 1



Nama : Arifah Shafyan  
NRP : 3322600003  
Tempat/Tanggal Lahir : Surabaya, 27 Mei 2004  
Agama : Islam  
Jenis Kelamin : Perempuan  
Program Studi : D4 Sains Data Terapan  
Departemen : Departemen Teknik Informatika dan Komputer  
Institusi : Politeknik Elektronika Negeri Surabaya  
Alamat : Jl Dukuh Setro 3 no 39, Surabaya, Jawa Timur  
Nomor Telepon : +6281554449827  
Email : [arifah.shafyan@gmail.com](mailto:arifah.shafyan@gmail.com)

#### **Riwayat Pendidikan:**

2022 – sekarang : Politeknik Elektronika Negeri Surabaya  
2019 – 2022 : SMA Negeri 9 Surabaya  
2016 – 2019 : SMP IT Al Uswah Surabaya  
2010 – 2016 : SD IT Al Uswah Surabaya

#### **Riwayat Pelatihan:**

1. Latihan Keterampilan Manajemen Mahasiswa Pra Tingkat Dasar (LKMM Pra- TD) 2022

2. Latihan Keterampilan Manajemen Mahasiswa Tingkat Dasar (LKMM TD) 2023

### **Pengalaman Organisasi**

1. Tim Robot ERISA 2023-2024



## Lampiran 2. Biodata Penulis 2



Nama : Puput Ayu Setiawati  
NRP : 3322600004  
Tempat/Tanggal Lahir : Sidoarjo, 02 Juli 2003  
Agama : Islam  
Jenis Kelamin : Perempuan  
Program Studi : D4 Sains Data Terapan  
Departemen : Departemen Teknik Informatika dan Komputer  
Institusi : Politeknik Elektronika Negeri Surabaya  
Alamat : Dusun Biting, Desa Suko RT.12 RW.03, Kec. Sidoarjo, Kab. Sidoarjo, Jawa Timur  
Nomor Telepon : +6289676345181  
Email : [puputayusetiawati2@gmail.com](mailto:puputayusetiawati2@gmail.com)

### Riwayat Pendidikan:

2022 – sekarang : Politeknik Elektronika Negeri Surabaya  
2019 – 2022 : SMA Negeri 4 Sidoarjo  
2016 – 2019 : SMP Negeri 4 Sidoarjo  
2010 – 2016 : SD Negeri Suko II Sidoarjo

### Riwayat Pelatihan:

1. Latihan Keterampilan Manajemen Mahasiswa Pra Tingkat Dasar (LKMM Pra- TD) 2022
2. Latihan Keterampilan Manajemen Mahasiswa Tingkat Dasar (LKMM TD) 2023

3. RevoU Mini Course Data Analytics 2024
4. RevoU Mini Course Data Analytics 2025

**Pengalaman Organisasi:**

1. Staff Ahli Departemen Kewirausahaan HIMIT 2024-2025

### Lampiran 3. Surat Pernyataan Penerimaan Kerja Praktik



**KEMENTERIAN KEUANGAN REPUBLIK INDONESIA**  
**DIREKTORAT JENDERAL PAJAK**  
**KANTOR WILAYAH DIREKTORAT JENDERAL PAJAK JAWA TIMUR I**  
**KANTOR PELAYANAN PAJAK PRATAMA SURABAYA GUBENG**  
JALAN SUMATERA NOMOR 22-24, SURABAYA 60281  
TELEPON (031) 5036962, 5036963, FAKSIMILE (031) 5031566, LAMAIL [www.pajak.go.id](http://www.pajak.go.id)  
LAYANAN INFORMASI DAN PENGADUAN KRISIS PAJAK (021) 1500200  
SUREL [pengaduan@pajak.go.id](mailto:pengaduan@pajak.go.id), [informasi@pajak.go.id](mailto:informasi@pajak.go.id)

Nomor : S-1168/KPP.1104/2024 13 Desember 2024  
Sifat : Biasa  
Lampiran : -  
Hal : Jawaban Permohonan Praktik Kerja Lapangan (PKL) Politeknik Elektronika Negeri Surabaya

Yth. Kepala Politeknik Elektronika Negeri Surabaya  
Jl. Raya ITS, Sukolilo, Surabaya

Sehubungan dengan surat Ketua Departemen Teknik Informatika dan Komputer nomor 948/PL14/PK.01.06/XI/2024 perihal Permohonan Kerja Paktik, dengan ini kami sampaikan hal-hal sebagai berikut :

1. Permohonan tempat praktik kerja lapangan yang Saudara ajukan sesuai surat Nomor 948/PL14/PK.01.06/XI/2024 tanggal 18 November 2024 dapat disetujui untuk mahasiswa atas nama:

| No. | Nama                | NRP        | Program Studi         |
|-----|---------------------|------------|-----------------------|
| 1   | Arifah Shafyan      | 3322600003 | D4 Sains Data Terapan |
| 2   | Puput Ayu Setiawati | 3322600004 | D4 Sains Data Terapan |

2. Mahasiswa yang bersangkutan akan ditempatkan pada Subbagian Umum dan Kepatuhan Internal atau seksi ini lingkungan KPP Pratama Surabaya Gubeng sesuai dengan kebutuhan organisasi.
3. Kegiatan praktik kerja lapangan dilaksanakan pada tanggal 13 Januari 2025 sampai dengan 09 Mei 2025 mulai pukul 08.00 sampai dengan pukul 16.00 WIB.
4. Setiap peserta praktik kerja lapangan harus mematuhi tata tertib yang berlaku di KPP Pratama Surabaya Gubeng dan ketentuan yang ditegaskan dalam Surat Edaran Direktur Jenderal Pajak Nomor: SE-23/PJ/2012 tanggal 25 April 2012 tentang Pemberian Izin Penelitian/Praktik Kerja Lapangan di Lingkungan Direktorat Jenderal Pajak, di antaranya:
  - a. peserta magang kerja hanya diperkenankan melakukan kegiatan/pekerjaan yang tidak berkaitan dengan data wajib pajak;
  - b. izin PKL berlaku selama satu periode;
  - c. segala biaya yang timbul untuk kegiatan dimaksud, menjadi beban sepenuhnya peserta magang.

Demikian disampaikan, atas perhatiannya diucapkan terima kasih.

Kepala Kantor Pelayanan Pajak  
Pratama Surabaya Gubeng



Divertifikasi secara elektronik  
Sakdun




## Lampiran 4. Sertifikat


|  |   |
|--|---|
|   |  |
| <b>SERTIFIKAT</b>  |   |
| KET-13/KPP.1104/KPP.110401/2025  |   |
| Diberikan kepada   |   |
| <b>Arifah Shafyan</b>  |   |
| Program D4 Sains Data Terapan – Politeknik Elektronika Negeri Surabaya   |   |
| Telah mengikuti program magang di Lingkungan Kementerian Keuangan, KPP Pratama Surabaya Gubeng, yang dilaksanakan pada 13 Januari s.d. 9 Mei 2025. |   |
| Surabaya, 9 Mei 2025   |   |
| a.n. Kepala Kantor Pelayanan Pajak Pratama Surabaya Gubeng<br>Kepala Subbagian Umum dan Kepatuhan Internal   |   |
|   |   |
| Ditandatangani secara elektronik<br><b>HIKMAH</b>  |   |




  

|  |   |
|--|---|
|   |  |
| <b>SERTIFIKAT</b>  |   |
| KET-12/KPP.1104/KPP.110401/2025  |   |
| Diberikan kepada   |   |
| <b>Puput Ayu Setiawati</b>   |   |
| Program D4 Sains Data Terapan – Politeknik Elektronika Negeri Surabaya   |   |
| Telah mengikuti program magang di Lingkungan Kementerian Keuangan, KPP Pratama Surabaya Gubeng, yang dilaksanakan pada 13 Januari s.d. 9 Mei 2025. |   |
| Surabaya, 9 Mei 2025   |   |
| a.n. Kepala Kantor Pelayanan Pajak Pratama Surabaya Gubeng<br>Kepala Subbagian Umum dan Kepatuhan Internal   |   |
|   |   |
| Ditandatangani secara elektronik<br><b>HIKMAH</b>  |   |


## Lampiran 5. Form Bimbingan KP


|   |  |
|---|--|
|  | <b>CHECKSHEET</b>                      |
|   | 11                                     |
|   | Politeknik Elektronika Negeri Surabaya |
|   | Monitoring KP                          |
|   | CS-BAK.11-Rev.0                        |
|   | 0                                      |






|                  |   |   |
|------------------|---|---|
| Nama             | Arifah Shafyan  |  |
| NRP              | 3322600003  |   |
| Jurusan / Kelas  | Sains Data Terapan / 3 SDT  |   |
| Dosen Pembimbing | Bapak Ahmad Syauqi Ahsan, S.Kom., MT.   |   |
| Judul Laporan KP | Klasifikasi Sentimen Masyarakat Terhadap Layanan KPP Pratama Surabaya Gubeng Berbasis Naive Bayes Dan SVM |   |

| MINGGU KE- | HARI/TANGGAL            | MATERI PEMBAHASAN   | PARAF PEMBIMBING  |
|------------|-------------------------|---|---|
| 15         | Kamis,<br>24 April 2025 | <ul style="list-style-type: none"> <li>- Identifikasi masalah / kebutuhan perusahaan</li> <li>- Penyesuaian fokus KP dengan tema laporan</li> <li>- Persetujuan dosen terhadap topik masalah dan pendekatan yg digunakan</li> <li>- Diskusi awal metode yang akan diterapkan</li> </ul>     |    |
| 18         | Kamis,<br>15 Mei 2025   | <ul style="list-style-type: none"> <li>- Review progress pengolahan data</li> <li>- Implementasi metode klasifikasi</li> <li>- Saran dosen terkait perbaikan akurasi</li> <li>- Diskusi kendala teknis</li> <li>- Rencana atau menyusun grafik, tabel, visualisasi dalam laporan</li> </ul> |    |
| 22         | Kamis,<br>12 Juni 2025  | <ul style="list-style-type: none"> <li>- Struktur laporan</li> <li>- Hasil klasifikasi dan interpretasi</li> <li>- Review dan saran dosen terhadap penulisan laporan, tata bahasa, format</li> </ul>  |  |

**PERHATIAN :** Lembar ini harus dibawa pada saat bimbingan & diparaf oleh dosen pembimbing

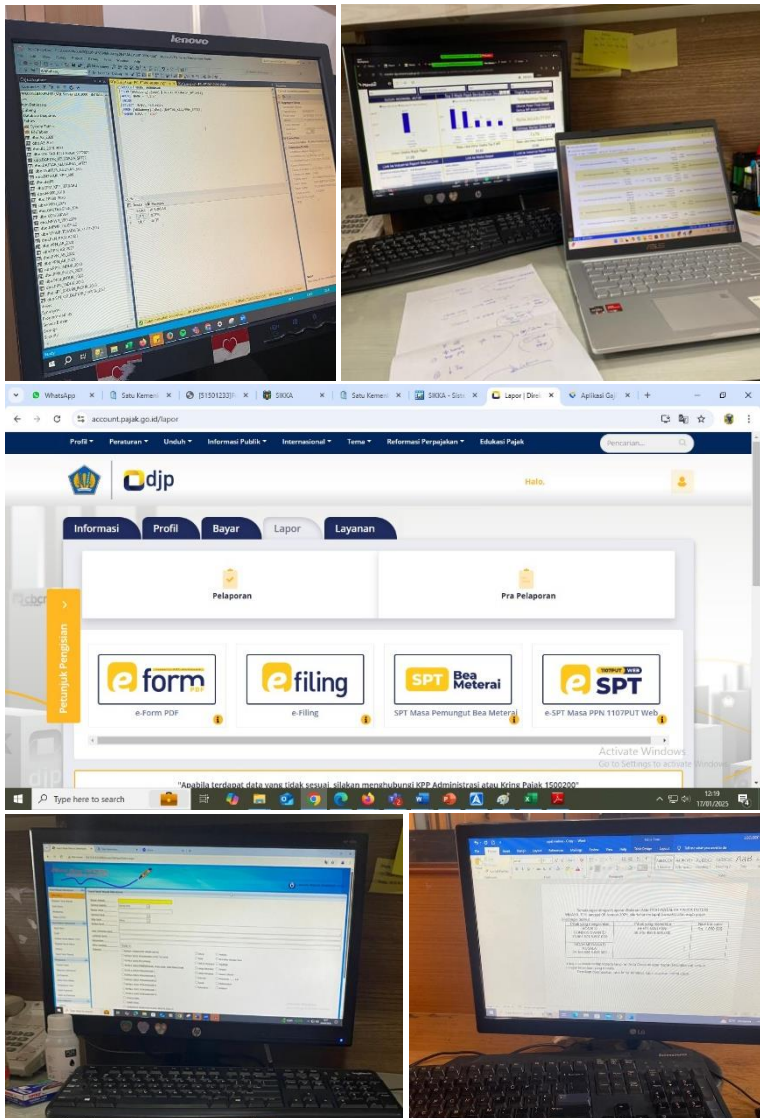
|   |  |
|---|--|
|  | <b>CHECKSHEET</b>                      |
|   | <b>11</b>                              |
|   | Politeknik Elektronika Negeri Surabaya |
|   | <b>Monitoring KP</b>                   |
|   | CS-BAK.11-Rev.0                        |
|   | 0                                      |

|                  |   |   |
|------------------|---|---|
| Nama             | Puput Ayu Setiawati   |  |
| NRP              | 3322600004  |   |
| Jurusan / Kelas  | D4 Sains Data Terapan / 3   |   |
| Dosen Pembimbing | Fitrah Maharani Humaira, M.Kom  |   |
| Judul Buku KP    | Klasifikasi Sentimen Masyarakat Terhadap Layanan KPP Pratama Surabaya Gubeng Berbasis Naive Bayes dan SVM |   |

| MINGGU KE- | HARI/TANGGAL             | MATERI PEMBAHASAN   | PARAF PEMBIMBING  |
|------------|--------------------------|---|---|
| 4          | Rabu<br>05 Februari 2025 | <ul style="list-style-type: none"> <li>Solusi Project dari KPP Pratama Surabaya Gubeng yang belum sesuai di saat awal-awal Kerja Praktik</li> </ul>                         |  |
| 14         | Selasa<br>15 April 2025  | <ul style="list-style-type: none"> <li>Diskusi mengenai penulisan buku KP</li> <li>Project yang diangkat menjadi buku KP</li> <li>Demo program sentimen analisis</li> </ul> |  |
| 17         | Kamis<br>08 Mei 2025     | <ul style="list-style-type: none"> <li>Persiapan berkas-berkas yang harus dilengkapi saat berakhirnya Kerja Praktik</li> </ul>  |  |
| 22         | Rabu<br>11 Juni 2025     | <ul style="list-style-type: none"> <li>Penyusunan buku KP</li> </ul>  |  |
| 22         | Jumat<br>13 Juni 2025    | <ul style="list-style-type: none"> <li>Penyusunan Buku KP</li> <li>Program sentimen analisis</li> </ul>   |  |
|            |                          |   |   |

**PERHATIAN** : Lembar ini harus dibawa pada saat bimbingan & diparaf oleh dosen pembimbing

## Lampiran 6. Dokumentasi



## **Lampiran 7. File Program**

Link Program:

[https://github.com/puputayus/KP\\_KPP-Pratama-Surabaya-Gubeng](https://github.com/puputayus/KP_KPP-Pratama-Surabaya-Gubeng)