

**ANALISIS SENTIMEN PADA ULASAN APLIKASI GOOGLE MAPS  
TERHADAP PELAYANAN BADAN PENYELENGGARA JAMINAN  
SOSIAL (BPJS) KESEHATAN SAMARINDA MENGGUNAKAN  
METODE K-NEAREST NEIGHBOR DENGAN FITUR EKSTRAKSI  
TF-IDF**

**SKRIPSI**

**Diajukan Oleh :**

**Ikhsan Nuttakwa Takbirata Ihram Nabawi**

**2011102441072**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR**

**JULI 2024**

**ANALISIS SENTIMEN PADA ULASAN APLIKASI GOOGLE MAPS  
TERHADAP PELAYANAN BADAN PENYELENGGARA JAMINAN  
SOSIAL (BPJS) KESEHATAN SAMARINDA MENGGUNAKAN  
METODE K-NEAREST NEIGHBOR DENGAN FITUR EKSTRAKSI  
TF-IDF**

**SKRIPSI**

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar Sarjana Teknik  
Informatika Fakultas Sains Dan Teknologi Universitas Muhammadiyah Kalimantan Timur

**Diajukan Oleh :**

**Ikhsan Nuttakwa Takbirata Ihram Nabawi**

**2011102441072**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR**

**JULI 2024**

## HALAMAN PERSETUJUAN

ANALISIS SENTIMEN PADA ULASAN APLIKASI GOOGLE MAPS TERHADAP  
PELAYANAN BADAN PENYELENGGARA JAMINAN SOSIAL (BPJS) KESEHATAN  
SAMARINDA MENGGUNAKAN METODE K-NEAREST NEIGHBOR DENGAN  
FITUR EKSTRAKSI TF-IDF

### SKRIPSI

Diajukan Oleh :

Ikhsan Nuttakwa Takbirata Ihram Nabawi  
2011102441072

Disetujui untuk diujikan

Pada tanggal 27 Juni 2024

Pembimbing



Rudiman, S.Kom., M.Sc

NIDN 1105068202

Mengetahui,

Koordinator Tugas Akhir/Skripsi/Tesis/Disertasi



Abdul Rahim, S.Kom., M.Cs

NIDN 0009047901

## HALAMAN PENGESAHAN

**ANALISIS SENTIMEN PADA ULASAN APLIKASI GOOGLE MAPS TERHADAP  
PELAYANAN BADAN PENYELENGGARA JAMINAN SOSIAL (BPJS)  
KESEHATAN SAMARINDA MENGGUNAKAN METODE K-NEAREST  
NEIGHBOR DENGAN FITUR EKSTRAKSI TF-IDF**

### Tugas akhir/skripsi

Diajukan oleh:

**Ikhsan Nuttakwa Takbirata Ihram Nabawi**

**2011102441072**

Diseminarkan dan Diujikan

Pada Tanggal 17 Juli 2024

Penguji I	Penguji II
 <b><u>Fendy Yulianto, S.Kom., M.Kom.</u></b> NIDN 1102079402	 <b><u>Rudiman, S.Kom, M.Sc</u></b> NIDN 1105068202

Mengetahui,

Ketua

**Program Studi Teknik Informatika**



**Arbansyah, S.Kom, M.TI**

NIDN. 1118019203

## PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Ikhsan Nuttakwa Takbirata Ihram Nabawi  
NIM : 2011102441072  
Program Studi : S1 Teknik Informatika  
Judul Penelitian : Analisis Sentimen Pada Ulasan Aplikasi Google Maps Terhadap Pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Samarinda Menggunakan Metode K – Nearest Neighbor Dengan Fitur Ekstraksi TF-IDF

Menyatakan bahwa Skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, dan bukan merupakan hasil plagiasi/falsifikasi/fabrikasi baik sebagian atau seluruhnya.

Atas pernyataan ini, saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apa bila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya saya ini.

Samarinda 30 Juni 2024

Yang membuat pernyataan



Ikhsan Nuttakwa Takbirata Ihram Nabawi

NIM: 2011102441072

## ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pelayanan BPJS Kesehatan Samarinda berdasarkan ulasan di aplikasi Google Maps. Metode yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor* (KNN) dengan fitur ekstraksi TF-IDF (*Term Frequency-Inverse Document Frequency*). Data yang digunakan terdiri dari 500 ulasan berbahasa Indonesia yang dikumpulkan melalui teknik *web scraping*. Setelah proses pengumpulan data, data dilabelling oleh seorang ahli (*expert*) dilakukan tahap pre-processing yang terdiri dari *case folding*, *cleaning*, *tokenizing*, *stop word removal*, dan *stemming*. Data kemudian diberi bobot menggunakan metode TF-IDF untuk mengidentifikasi kata-kata penting. Pengujian dilakukan dengan menggunakan rasio data latih dan data uji sebesar 70:30 serta nilai  $k = 5$ . Hasilnya menunjukkan bahwa metode KNN mampu mengklasifikasikan sentimen positif dan negatif dengan tingkat akurasi sebesar 93,3%. Analisis ini memberikan gambaran tentang kualitas pelayanan BPJS Kesehatan di Samarinda dan dapat digunakan sebagai dasar untuk perbaikan layanan. Selain itu, penelitian ini juga memberikan kontribusi dalam penggunaan KNN dan TF-IDF untuk analisis sentimen, membuka peluang untuk penelitian lebih lanjut di bidang ini.

**Kata kunci:** Analisis Sentimen, BPJS Kesehatan, K-Nearest Neighbor

## **ABSTRACT**

*This study aims to analyze public sentiment towards the services of BPJS Kesehatan Samarinda based on reviews on the Google Maps application. The method used in this research is K-Nearest Neighbor (KNN) with TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction. The data used consists of 500 Indonesian-language reviews collected through web scraping techniques. After the data collection process, the data was labeled by an expert, and then a pre-processing stage was carried out, including case folding, cleaning, tokenizing, stop word removal, and stemming. The data was then weighted using the TF-IDF method to identify important words. The testing was conducted using a training and testing data ratio of 70:30 and a k value of 5. The results showed that the KNN method was able to classify positive and negative sentiments with an accuracy rate of 93.3%. This analysis provides an overview of the service quality of BPJS Kesehatan in Samarinda and can be used as a basis for service improvements. Additionally, this research contributes to the use of KNN and TF-IDF for sentiment analysis, opening opportunities for further research in this field.*

**Keywords:** *Sentiment Analysis, BPJS Kesehatan, K-Nearest Neighbor*

## PRAKATA

Puji dan syukur saya panjatkan kepada Allah SWT. Atas ridho-Nya penulis dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang penulis ajukan adalah “Analisis Sentimen Pada Ulasan Aplikasi Google Maps Terhadap Pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Samarinda Menggunakan Metode *K – Nearest Neighbor* Dengan Fitur Ekstraksi TF-IDF”

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Skripsi di Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur. Tidak dapat disangka bahwa butuh usaha yang keras dalam penyelesaian pengerjaan skripsi ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling penulis yang mendukung dan membantu. Terima kasih penulis sampaikan kepada:

1. Bapak Dr. Muhammad Musiyam, M.T. Selaku Rektor Universitas Muhammadiyah Kalimantan Timur yang telah memberi kesempatan pada peneliti untuk menempuh pendidikan di Universitas Muhammadiyah Kalimantan Timur
2. Bapak Rudiman, M.Sc. Selaku Dosen Pembimbing penulis yang telah membagi ilmu, bimbingan dan arahan selama proses perkuliahan.
3. Orang tua dan saudari penulis yang selalu mendukung, memotivasi dan mendoakan penulis sehingga proses pengerjaan skripsi ini selesai sampai akhir.
4. Pacar penulis yang bernama Nur Adelia Raihanda yang telah membantu penulis, menjadi *support system* dan mendoakan penulis sampai akhirnya penulis bisa mengerjakan skripsi ini sampai selesai.

5. Sahabat penulis, Khoirul Huda, Rayhan Elfansyah, Ipan Hasmadi, Farhat Jundullah, Zulkarnaen, dan anggota kos lainnya terimakasih atas dukungan dan semangat yang diberikan kepada penulis selama ini.
6. Teman satu bimbingan dan teman kuliah peneliti yang bernama Takhta Perlawanan, Ricky, Andi Nur Halim, Aditya dll yang sudah membantu peneliti dan mensupport peneliti, bersama-sama menyelesaikan skripsi ini sampai akhir.

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah SWT. Dan akhirnya penulis menyadari bahwa skripsi ini tidak luput dari berbagai kekurangan, karena keterbatasan ilmu yang saya miliki. Untuk itu dengan kerendahan hati saya mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini,

Samarinda, 30 Juni 2024



Ikhsan Nuttakwa Takbirata Ihram Nabawi

# DAFTAR ISI

	<b>Halaman</b>
HALAMAN PERSETUJUAN .....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN KEASLIAN PENELITIAN .....	iii
ABSTRAK .....	iv
<i>ABTRACT</i> .....	v
PRAKATA .....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN .....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
BAB II METODELOGI PENELITIAN .....	5
2.1. Objek Penelitian .....	5
2.2. Alat dan Bahan .....	5
2.3. Prosedur Penelitian .....	5
2.3.1. Pengumpulan Data .....	6
2.3.2. <i>Labeling Data</i> .....	7
2.3.3. <i>Pre-Processing</i> .....	8
2.3.4. Pembobotan Kata (TF-IDF) .....	9
2.3.5. <i>Split Data</i> .....	10
2.3.6. Klasifikasi.....	10
2.3.7. <i>K-Nearest Neighbors</i> (KNN) .....	10
2.3.8. Evaluasi .....	11
2.4 Jadwal Penelitian .....	12
BAB III HASIL DAN PEMBAHASAN.....	14
3.1 Pengumpulan Data.....	14
3.2 Labelling Data .....	15

3.3	<i>Pre-Processing</i> .....	16
3.3.1	<i>Case Folding</i> .....	16
3.3.2	<i>Cleaning</i> .....	17
3.3.3	<i>Tokenizing</i> .....	18
3.3.4	<i>Stopword Removal</i> .....	18
3.3.5	<i>Stemming</i> .....	19
3.4	Pembobotan Kata (TF-IDF).....	19
3.5	Visualisasi.....	21
3.6	Evaluasi .....	22
BAB IV PENUTUP .....		28
4.1	Simpulan.....	28
4.2	Implikasi .....	28
DAFTAR RUJUKAN .....		30
LAMPIRAN .....		32
RIWAYAT HIDUP .....		47

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 2.1 Jadwal Penelitian.....	13
Tabel 3.1 Pengumpulan Data .....	14
Tabel 3.2 Labeling Data .....	15
Tabel 3.3 Hasil Data Case Folding.....	17
Tabel 3.4 Hasil Data Cleaning .....	17
Tabel 3.5 Hasil Data Tokenizing.....	18
Tabel 3.6 Hasil Data Stopword Removal.....	18
Tabel 3.7 Hasil Data Stemming .....	19
Tabel 3.8 Pembagian Data KNN.....	22
Tabel 3.9 Hasil Data Uji.....	26

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 2.1 Alur Penelitian.....	6
Gambar 2.2 Hasil Crawling Data .....	7
Gambar 3.1 Visualisasi Sentimen .....	16
Gambar 3.2 Pembobotan Kata (TF-IDF) .....	20
Gambar 3.3 Scoring TF .....	21
Gambar 3.4 WordCloud Positif.....	21
Gambar 3.5 WordCloud Negatif.....	22
Gambar 3.6 Hasil Akurasi Nilai K.....	24
Gambar 3.7 Plot Akurasi Nilai K.....	25
Gambar 3.8 Hasil Akurasi Metode KNN.....	26
Gambar 3.9 Hasil Confusion Matrix .....	26

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
L1 CV Expert Labeling.....	32
L2 Surat Izin Penelitian.....	34
L3 Pengumpulan Data.....	35
L4 Install Library .....	37
L5 Membaca Dataset.....	37
L6 Cek Informasi Dataset .....	38
L7 Cek Jumlah Sentimen Positif dan Negatif.....	38
L8 Persentase Sentimen .....	38
L9 Membaca Dataset Kolom Review dan Sentimen .....	38
L10 Pre-Processing Data.....	39
L11 Pembobotan Kata TF-IDF .....	40
L12 Split Data dan Evaluasi Confusion Matrix Metode KNN .....	41
L13 Hasil Akurasi Pada Nilai K.....	41
L14 Plot Akurasi Nilai K .....	42
L15 Visualisasi Wordcloud Positif dan Negatif.....	42
L16 Kartu Kendali Bimbingan.....	43

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Badan Penyelenggara Jaminan Sosial (BPJS) merupakan suatu lembaga yang didirikan untuk mengelola program jaminan sosial, dengan fokus utama pada aspek kesehatan dan ketenagakerjaan, Jaminan Kesehatan Nasional (JKN) telah diselenggarakan sejak tahun 2014, semua penduduk Indonesia wajib menjadi peserta BPJS termasuk orang asing yang telah bekerja di Indonesia (Linda et al., 2020). Salah satu tujuan dan manfaat dari BPJS adalah untuk memberikan layanan yang baik bagi peserta dalam bidang kesehatan, dan juga mempermudah akses masyarakat dalam mendapatkan layanan dan kini manfaatnya semakin meluas dengan memfasilitasi masyarakat dalam mendapatkan pelayanan publik di berbagai sektor dan industri (Annisa Medina Sari, 2023).

Berdasarkan data dari [bpjs-kesehatan.go.id](http://bpjs-kesehatan.go.id) sampai dengan 31 Desember 2023 jumlah peserta BPJS Kesehatan sebanyak 267 juta penduduk Indonesia, sedangkan pada tahun 2022 total peserta BPJS Kesehatan mencapai 248 juta orang yang artinya peserta BPJS Kesehatan Mengalami Kenaikan Mencapai 19 juta orang (BPJS Kesehatan, n.d.). Untuk di Kota Samarinda total penduduk sebanyak 838.935 jiwa, jumlah peserta JKN sebanyak 837.890 jiwa, hal ini menunjukkan bahwa sebagian besar orang Indonesia menggunakan BPJS Kesehatan (Antara Kaltim, 2023). Seiring berjalannya waktu semakin banyak orang di Indonesia yang menggunakan dan memanfaatkan fasilitas yang disediakan yaitu layanan BPJS, dikarenakan hal tersebut maka banyak timbul opini serta keluhan yang diungkapkan dari peserta mengenai layanan BPJS. (Puspita & Widodo, 2021).

Terdapat berbagai opini negatif dan positif mengenai layanan BPJS di seluruh Indonesia yang ditemukan melalui pengamatan berbagai media online, seperti masyarakat mengeluhkan antrian panjang hingga 5-6 jam (Khadijah Nur Azizah, 2021) serta tanggapan positif seperti biaya pelayanan kesehatan tidak menjadi bebannya lagi berkat bantuan BPJS Kesehatan (CNN Indonesia, 2021). Pada pelayanan BPJS Kesehatan di kota Samarinda juga terdapat opini mengenai layanan BPJS seperti proses panjang yang menyulitkan pengguna kartu yang sedang sakit (IDN Times Kaltim, 2019). Dari berbagai keluhan dan opini tersebut peneliti tertarik untuk menganalisis mengenai bagaimana sentimen masyarakat terhadap pelayanan yang disediakan oleh BPJS Kesehatan di Samarinda.

Dalam penelitian ini metode analisis sentimen digunakan untuk menganalisis pendapat atau opini berdasarkan ulasan pengguna Google Maps. Analisis sentimen adalah suatu metode untuk menganalisis, memproses, dan mengubah data berupa teks yang terkait dengan suatu objek seperti layanan, produk, figur publik, peristiwa, atau topik tertentu. Analisis sentimen dapat dianggap sebagai penggalan opini yang berfokus pada penelitian tentang opini positif dan negatif (Parasian Doloksaribu & Samuel, 2022). Analisis sentimen juga merupakan proses pengelompokan dari teks yang terdapat dalam kalimat atau dokumen dengan tujuan untuk menentukan apakah pendapat yang terungkap dalam kalimat atau dokumen tersebut bersifat positif atau negatif (Amrullah et al., 2020).

Terdapat beberapa penelitian sebelumnya yang pertama dilakukan oleh Rani Puspita dan Agus Widodo yang berjudul "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS". Dengan metode KNN peneliti berhasil mendapatkan akurasi 95.58%, Decision Tree 96.13%, dan Naïve Bayes 89,14% (Puspita & Widodo, 2021). Penelitian kedua oleh Al Khadafi yang berjudul "Penerapan metode Naïve Bayes Classifier dan Lexicon Based untuk analisis sentimen cyberbullying pada BPJS", peneliti menggunakan fitur TF-IDF untuk pembobotan kata berhasil menganalisis sentimen dengan

tingkat akurasi untuk Naïve Bayes 80% sedangkan Lexicon Based 22% (Al Khadafi et al., 2022).

Peneliti menggunakan metode *K – Nearest Neighbor* untuk melihat berapa akurasi yang dihasilkan dari metode tersebut. Metode *K-Nearest Neighbor* (KNN) digunakan untuk proses klasifikasi dalam sebuah penelitian karena memiliki kesederhanaan yang berbasis pada pendekatan pembobotan yang sederhana. Selain itu, KNN juga dikenal karena kemudahan dalam implementasi, adaptasi, dan proses pembelajaran. Keunggulan KNN yaitu mudah diimplementasikan karena cukup mendefinisikan fungsi untuk menghitung jarak antar kelas (Salim & Mayary, 2020).

Berdasarkan uraian di atas, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini publik mengenai pelayanan BPJS Kesehatan di Samarinda. Opini-opini tersebut akan dikelompokkan menjadi sentimen positif dan negatif dengan menggunakan metode KNN dan fitur TF-IDF untuk meningkatkan efisiensi proses klasifikasi. Tujuan utama dari penelitian ini adalah menggunakan hasil analisis sentimen untuk mengevaluasi kualitas layanan BPJS Kesehatan di Samarinda saat ini. Hasil yang didominasi oleh opini positif akan menunjukkan kualitas layanan yang baik yang dapat ditingkatkan lebih lanjut, sementara kecenderungan opini negatif akan mengidentifikasi aspek-aspek yang perlu perbaikan dalam layanan tersebut.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang diatas, maka rumusan masalah :

1. Bagaimana hasil pembobotan TF-IDF dalam mengidentifikasi kata-kata penting dalam ulasan BPJS Kesehatan Samarinda di Google Maps?

2. Bagaimana hasil akurasi menggunakan metode *K-Nearest Neighbor* dalam mengklasifikasikan opini pengguna Google Maps terhadap layanan BPJS Kesehatan di Samarinda?

### **1.3 Tujuan Penelitian**

Penelitian ini memiliki beberapa tujuan :

1. Mengetahui hasil pembobotan TF-IDF dalam mengidentifikasi kata-kata penting dalam ulasan BPJS Kesehatan Samarinda di Google Maps?
2. Mengetahui akurasi menggunakan metode *K-Nearest Neighbor* dalam mengklasifikasikan opini pengguna Google Maps terhadap layanan BPJS Kesehatan di Samarinda.

### **1.4 Manfaat Penelitian**

Terdapat beberapa manfaat dari penelitian ini :

1. Memberikan data yang berguna bagi BPJS Kesehatan Samarinda untuk meningkatkan kualitas pelayanan berdasarkan masukan yang diterima dari pengguna.
2. Memberikan informasi kepada masyarakat tentang sentimen pengguna terhadap pelayanan BPJS Kesehatan di Samarinda, sehingga dapat mengetahui pelayanannya baik atau kurang baik.
3. Memberikan pengalaman dalam melakukan penelitian yang melibatkan analisis sentimen dan penggunaan metode KNN dengan fitur ekstraksi TF-IDF.

## **BAB II**

### **METODELOGI PENELITIAN**

#### **2.1. Objek Penelitian**

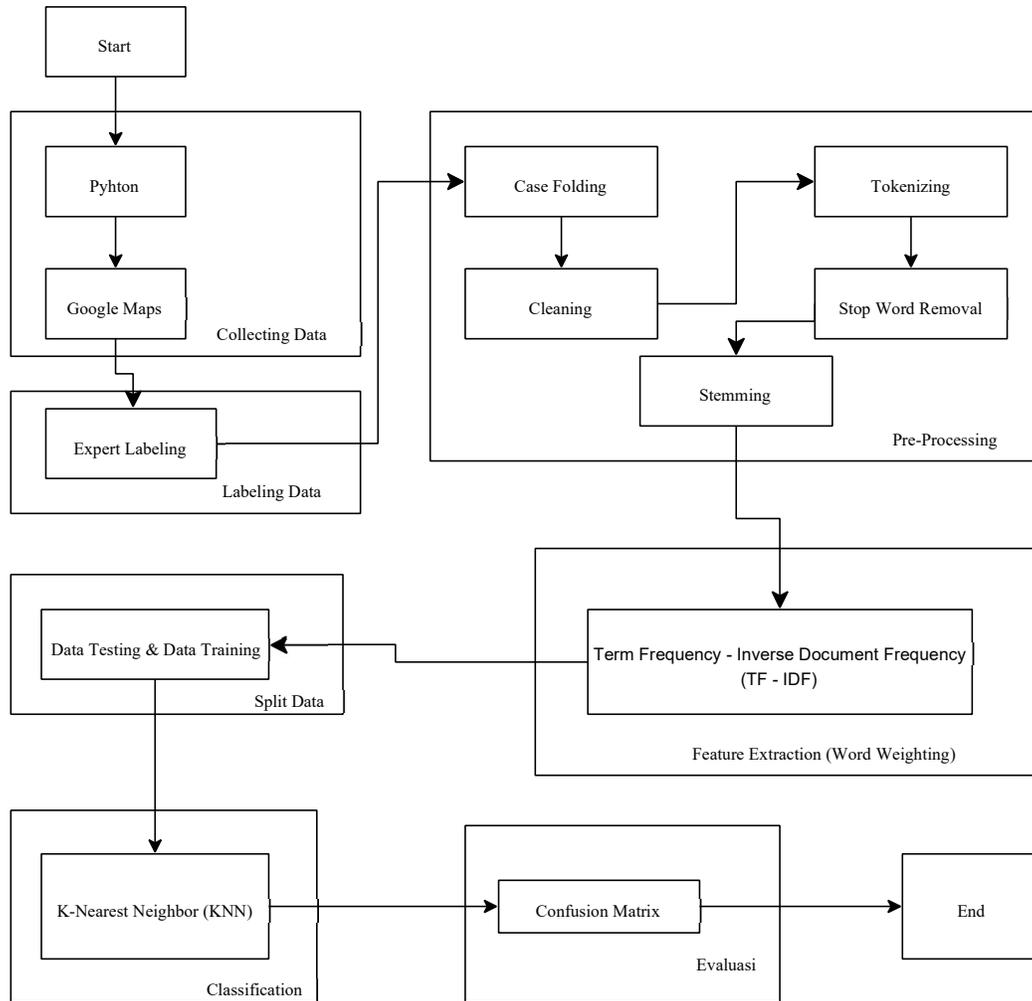
Objek penelitian dalam skripsi ini adalah ulasan pengguna Google Maps terhadap layanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Samarinda. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat yang menggunakan layanan BPJS Kesehatan tersebut dengan menggunakan algoritma *K-Nearest Neighbor* (KNN). Proses penelitian melibatkan beberapa tahap, yaitu pengumpulan data ulasan melalui teknik *crawling* menggunakan *library BeautifulSoup4*, labelling data, *pre-processing* data, pembobotan kata dengan TF-IDF, pembagian data menjadi data *training dan testing*, serta klasifikasi dan evaluasi hasil. Data yang digunakan berjumlah 500 ulasan berbahasa Indonesia yang kemudian diklasifikasikan oleh seorang ahli bahasa (*expert*) menjadi sentimen positif dan negatif untuk mengevaluasi kualitas layanan BPJS Kesehatan di Samarinda.

#### **2.2. Alat dan Bahan**

Dalam penelitian ini, alat dan bahan yang digunakan meliputi perangkat keras berupa laptop dengan prosesor Intel Core i5-1035G1, RAM 12GB, dan penyimpanan SSD 256GB. Untuk perangkat lunak digunakan Google Colab versi 1.0.0 (<https://colab.research.google.com/>) dengan *Python* versi 3.8.10. *Library Python* yang digunakan mencakup *Beautifulsoup4* versi 4.12.3, *Pandas* versi 2.0.3, *NumPy* versi 1.25.2, *NLTK* versi 3.8.1, *Scikit-learn* versi 1.2.2, *Matplotlib* versi 3.7.1, *Sastrawi* versi 1.0.1.

#### **2.3. Prosedur Penelitian**

Prosedur penelitian adalah langkah-langkah yang diambil oleh peneliti untuk menghimpun data atau informasi untuk kemudian dianalisis secara ilmiah. Terdapat 8 tahapan yang menjadi dasar bagi sebuah penelitian yang ditampilkan pada Gambar 2.1.



**Gambar 2.1** Alur Penelitian

Dalam penelitian ini, algoritma yang digunakan adalah *K-Nearest Neighbor* (KNN) untuk menganalisis sentimen terhadap pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Samarinda. Proses dilakukan melalui beberapa tahap : Pengumpulan data, Labeling Data oleh ahli bahasa (*expert*), *Pre-processing*, Pembobotan kata (TF-IDF), Pembagian data (*Split data*), Klasifikasi, dan Evaluasi.

### 2.3.1. Pengumpulan Data

Dalam penelitian ini pengambilan data dilakukan melalui proses *crawling* data ulasan Google Maps dengan memanfaatkan *library BeautifulSoup4* yang diimplementasikan dalam bahasa pemrograman Python. *BeautifulSoup4* digunakan sebagai alat untuk mengumpulkan

data ulasan dari Google berdasarkan identifikasi unik sebuah tempat (*feature\_id*) dan menganalisis informasi yang terkandung dalam ulasan tersebut, seperti rating dan sentimen pengguna. Dengan menggunakan *BeautifulSoup4*, peneliti dapat mengekstrak data ulasan yang relevan, mengidentifikasi sentimen dari ulasan (positif atau negatif), dan mengumpulkan informasi penting lainnya dari ulasan pengguna. (Amanny Ulfah Nabiylah Ramadhanty, 2023). Hasil *crawling* data pada Gambar 2.2.

```
27 token is None.  
    break # Hentikan loop jika token berikutnya kosong  
  
# Gabungkan semua DataFrame menjadi satu  
combined_df = pd.concat(dfs, ignore_index=True)  
  
# Simpan DataFrame gabungan ke dalam file Excel  
combined_df.to_excel("review_data.xlsx", index=False)  
  
print("Total data yang sudah didapat:", total_data)  
print("Data telah disimpan dalam review_data.xlsx")  
  
get_reviews_data(feature_id)  
  
Loop ke-33:  
-----  
Loop ke-34:  
-----  
Loop ke-35:  
-----  
Loop ke-36:  
-----  
Loop ke-37:  
-----  
Loop ke-38:  
-----
```

**Gambar 2.2** Hasil *Crawling* Data

Data yang berhasil diekstrak kemudian disimpan dalam format file Excel, memudahkan analisis dan pengolahan data lebih lanjut.

### 2.3.2. *Labeling Data*

Dalam proses melakukan klasifikasi teks pada data komentar sebagai bagian dari skripsi tugas akhir, peneliti membutuhkan ahli bahasa (*expert*) yang memiliki pengalaman dalam pelabelan data dan memiliki pengetahuan mendalam tentang bahasa Indonesia. Untuk itu,

peneliti mengajukan permintaan pada website project.co.id untuk mencari tenaga ahli yang sesuai dengan kriteria tersebut. Dalam pengajuan tersebut, peneliti menjelaskan bahwa dibutuhkan lulusan dari jurusan Bahasa Indonesia yang saat ini bekerja dalam bidang terkait seperti guru/dosen bahasa Indonesia, penulis, atau ahli bahasa. Calon tenaga ahli diminta untuk memasukkan penawaran dengan mencantumkan pekerjaan saat ini, pengalaman yang relevan dengan bahasa Indonesia, serta gelar akademik yang dimiliki.

### 2.3.3. *Pre-Processing*

*Preprocessing* adalah langkah awal dalam klasifikasi teks yang bertujuan untuk menyiapkan data teks sebelum digunakan dalam proses lanjutan. Pada tahap ini, data teks disesuaikan agar menghasilkan informasi yang lebih berkualitas dan siap untuk digunakan dalam langkah-langkah berikutnya (Khairunnisa et al., 2021). Langkah-langkah dilakukan dalam tahap *pre-processing* (Ratih Puspitasari et al., 2023) :

- 1) *Case folding* adalah proses yang mengubah huruf kapital dalam dokumen menjadi huruf kecil sebagai standar.
- 2) *Cleaning* adalah tahap di mana karakter yang tidak diperlukan seperti URL, tanda @, #, https:, RT (*Retweet*), angka, simbol, dan emotikon dihapus dari dokumen.
- 3) *Tokenizing* adalah proses memecah kalimat dalam dokumen menjadi kata-kata, di mana tanda baca, simbol, dan karakter bacaan yang tidak bernilai dihilangkan.
- 4) *Stopword Removal* adalah langkah untuk menghilangkan kata-kata yang memiliki tingkat informasi rendah. Ini dilakukan jika kata-kata tersebut termasuk dalam kategori umum dan tidak signifikan seperti kata penghubung, waktu, dan sejenisnya.
- 5) *Stemming* adalah proses menghilangkan awalan dan akhiran kata sehingga menjadi bentuk dasarnya. Stemming sering menggunakan library Sastrawi dalam bahasa pemrograman Python.

#### 2.3.4. Pembobotan Kata (TF-IDF)

Pada tahap ini, dilakukan pembobotan kata dalam teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF adalah Salah satu metode pembobotan yang menggabungkan frekuensi term (TF) dan frekuensi dokumen terbalik (IDF) adalah TF-IDF. *Term frequency* (TF) mengukur seberapa sering sebuah kata muncul dalam suatu dokumen, sedangkan *inverse document frequency* (IDF) mengukur seberapa banyak dokumen dalam keseluruhan korpus yang mengandung kata tertentu (Assidyk et al., 2020). Tujuan dari langkah ini adalah untuk menentukan bobot kata dalam sebuah dokumen atau seberapa sering kata tersebut muncul dalam dokumen tersebut. (Fadiyah Basar et al., 2022).

Penggunaan TF dapat menggunakan rumus pada Persamaan (2.1).

$$tf_{ij} = \frac{f_d(i)}{\max f_d(j)} \quad (2.1)$$

TF menunjukkan dokumen (d) seberapa banyak kata (t) yang muncul. Dan terkait untuk rumus IDF bisa dilihat pada Persamaan (2.2).

$$idf_t = \log \frac{N}{df_t} \quad (2.2)$$

N melambangkan jumlah kata dalam teks, df adalah jumlah teks yang memiliki kata t. Dengan menggabungkan TF dan IDF dalam pengerjaan dapat membantu meningkatkan performa. Terkait rumus pembobotan TF-IDF bisa dilihat pada Persamaan (2.3).

$$W_{t,d} = tf_{d,t} \times idf_t \quad (2.3)$$

Keterangan :

t = Kata kunci, term

d = Dokumen

W<sub>d,t</sub> = Bobot d terhadap t

Tf = Banyaknya t (kata) yang dicari dalam dokumen

Idf = Banyak t kebalikan dari kata yang dicari

### **2.3.5. Split Data**

Pada tahap ini, proses *split data* dilakukan untuk membagi dataset yang digunakan dalam penelitian menjadi dua bagian: data latih (*training*) dan data uji (*testing*). Data latih digunakan untuk melatih algoritma, sementara data uji digunakan untuk mengevaluasi kinerja algoritma tersebut (Putri et al., 2023). Dalam penelitian ini, data dibagi dengan rasio 90:10, 80:20, dan 70:30 untuk mengevaluasi pengaruh berbagai perbandingan rasio terhadap kinerja model. Hasil dari pembagian data ini akan menunjukkan bagaimana variasi rasio tersebut dapat mempengaruhi tingkat akurasi model dalam mengklasifikasikan sentimen.

### **2.3.6. Klasifikasi**

Klasifikasi merupakan salah satu tahap penting dalam text mining yang bertujuan untuk mengelompokkan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu. Proses ini melibatkan penggunaan teknik yang melihat variabel dari kelompok data yang sudah ada untuk menentukan aturan pengelompokan. Dengan mempelajari pola dari data yang sudah diberi label, klasifikasi memungkinkan kita untuk memprediksi kelas dari suatu objek yang belum diketahui sebelumnya (Azzahra Nasution et al., 2019).

### **2.3.7. K-Nearest Neighbors (KNN)**

Klasifikasi KNN merupakan metode non-parametrik sederhana yang digunakan untuk klasifikasi. Meskipun algoritma ini sederhana, kinerjanya sangat baik dan menjadi metode tolok ukur yang penting. Klasifikasi KNN membutuhkan metrik dan integer positif ( $K$ ). Aturan KNN memegang posisi sampel pelatihan beserta kelas mereka. Ketika menghadapi data masuk baru, tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan nilai atribut dan data latih yang ada (Putra et al., 2022).

Langkah-langkah klasifikasi algoritma KNN:

- a) Tentukan parameter nilai  $k$  = banyaknya jumlah tetangga terdekat.
- b) Hitung jarak antara data *training* dan data *testing*, rumusnya pada Persamaan (2.4).

$$euc = \sqrt{(\sum_{i=1}^n (p_i - q_i)^2)} \quad (2.4)$$

Keterangan :

$p_i$  = sample data / *data training*

$q_i$  = data uji / *data testing*

$I$  = variabel data

$n$  = dimensi data

- c) Urutkan jarak-jarak tersebut dan tetapkan tetangga terdekat berdasarkan jarak minimum hingga ke- $k$ .
- d) Periksa kelas dari tetangga terdekat.
- e) Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi untuk data baru.

### 2.3.8. Evaluasi

Pada tahap Evaluasi, dilakukan analisis Akurasi dengan menggunakan confusion matrix pada dokumen yang telah diklasifikasikan oleh algoritma *K-Nearest Neighbor* (KNN). Ketika mengevaluasi kinerja menggunakan *confusion matrix*, ada empat istilah yang mencerminkan hasil klasifikasi. Istilah-istilah tersebut meliputi *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Negative* (TN) mewakili jumlah data negatif yang berhasil teridentifikasi secara benar, sementara *False Positive* (FP) adalah data negatif yang salah teridentifikasi sebagai data positif. (Rifa et al., 2023).

## 1. Akurasi

Akurasi dapat dijelaskan sebagai ukuran seberapa dekat nilai prediksi dengan nilai sebenarnya. Semakin tinggi akurasi, semakin baik proses klasifikasi tersebut.

Rumus untuk menghitung akurasi ditunjukkan dalam Persamaan (2.5).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.5)$$

Dimana:

TP = *True Positif*

TN = *True Negatif*

FP = *False Positif*

FN = *False Negatif*

Pada rumus di atas, cara menentukan akurasi dari sebuah data dapat dilihat dengan menghitung jumlah prediksi yang benar (*True Positive dan True Negative*) dan membaginya dengan jumlah total prediksi yang dilakukan.

## 2.4 Jadwal Penelitian

Penelitian ini akan diawali dengan penentuan judul, identifikasi masalah, studi literatur, rancangan metode, pemilihan studi kasus, dan penyusunan proposal. Setelah tahap pra-penelitian selesai, penelitian akan melanjutkan ke pengumpulan data menggunakan metode *crawling* di google maps. Data yang terkumpul akan dilabeli oleh ahli Bahasa (*expert*) dan diproses sebelum pembobotan kata dengan metode TF-IDF. Setelahnya data akan dibagi menjadi data pelatihan dan pengujian sebelum dilakukan klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Tahap akhir adalah evaluasi dan analisis hasil untuk mengevaluasi efektivitas metode yang digunakan. Setelah selesai, akan dilakukan penyusunan laporan dan presentasi seminar untuk memperkenalkan hasil penelitian. Penelitian akan dilaksanakan sesuai jadwal terperinci pada Tabel 2.1.

**Tabel 2.1** Jadwal Penelitian

No	Kegiatan	Bulan/2024					
		Feb	Mar	Apr	Mei	Juni	Juli
Tahap Pra Penelitian							
1	Menentukan Judul	■					
2	Identifikasi Masalah	■					
3	Studi Literatur		■				
4	Rancangan Metode		■				
5	Pemilihan Studi Kasus		■				
6	Menyusun Proposal	■	■	■			
7	Review Desk Simpel				■		
Tahap Penelitian							
1	Pengumpulan Data ( <i>Crawling</i> )			■	■		
2	Labelling Data				■		
3	Pre-Processing Data				■		
4	Pembobotan Kata (TF-IDF)					■	
5	Split Data					■	
6	Klasifikasi K – Nearest Neighbor					■	
7	Evaluasi dan Analisis Hasil					■	
Tahap Akhir Penelitian							
1	Penyusunan Laporan					■	■
2	Seminar Hasil						■

## BAB III HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Pada penelitian ini, digunakan data ulasan terkait BPJS di kota Samarinda yang diambil dari Google Maps. Data yang berhasil dikumpulkan sebanyak 500 ulasan, dan disimpan dalam format CSV. Proses pengumpulan data dilakukan dengan metode *crawling* selama 5 menit pada tanggal 7 Juni 2024, dari pukul 21:20 WITA hingga pukul 21:25 WITA. Hasil pengumpulan data pada Tabel 3.1.

**Tabel 3.1** Pengumpulan Data

<b>No.</b>	<b>Name</b>	<b>Rating</b>	<b>Review</b>	<b>Time</b>
1	Hasan Fajar	5	Pelayanan Luarbiasa, Cepat, tepat.. ramah	Jumat 31-05-2024
2	Eka Kurnia Bawai	5	Saya senang dengan pelayanannya bagus dan petugasnya rama dalam menjelaskannya.	Jumat 31-05-2024
3	Nei San	5	Alhamdulillah pelayanan bagus. Tempatnya dingin dan mudah ditemukan. Cs, security dan yg bertugas yg lain sangat ramah. Sopan dan menjelaskan dgn baik. Terus pertahankan. Terima kasih	Rabu 08-05-2024
4	Yogi Ahmad	5	Penghargaan untuk seluruh aspek disini, fasilitas ok dan staffnya ramah. Pilihannya pertahankan atau tingkatkan.	Selasa 09-01-2024

<i>No.</i>	<i>Name</i>	<i>Rating</i>	<i>Review</i>	<i>Time</i>
5	Atifa Raihana	5	Pokok e de best sekali. Pelayanan ramah serta cepat.... makasih bpjs kesehatan telah membantu saya....	Jumat 31-05-2024
Total				500

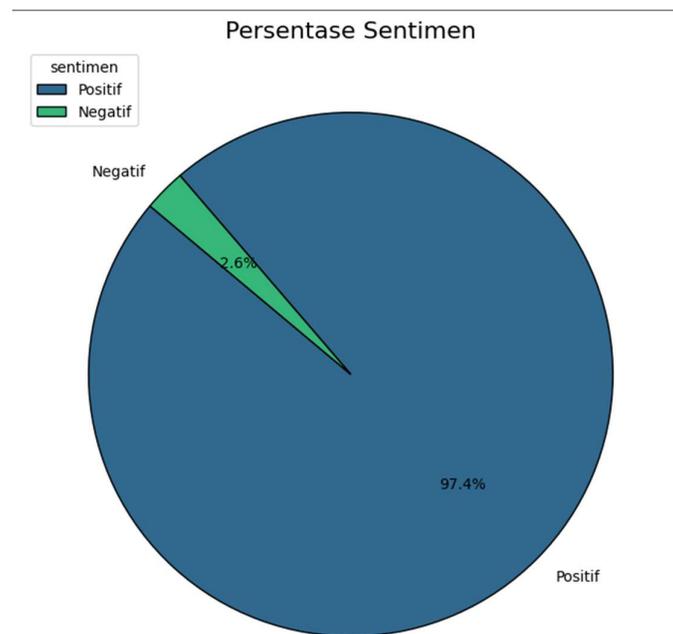
### 3.2 Labelling Data

Pelabelan data dilakukan oleh ahli bahasa (*expert*) untuk 500 data ulasan pada Google Maps. Data ulasan pada penelitian ini akan dikategorikan ke dalam dua kelompok utama yaitu positif, dan negatif. Dari data yang telah dilabeli, terlihat bahwa sebagian besar data menunjukkan sentimen positif. Terdapat dua kelompok ulasan setelah diberi label pada Tabel 3.2.

**Tabel 3.2** *Labeling Data*

No	<i>Review</i>	Sentimen
1	Pelayanan Luarbiasa, Cepat, tepat.. ramah	Positif
2	Saya senang dengan pelayanannya bagus dan petugasnya rama dalam menjelaskannya.	Positif
3	Alhamdulillah pelayanan bagus. Tempatnya dingin dan mudah ditemukan. Cs, security dan yg bertugas yg lain sangat ramah. Sopan dan menjelaskan dgn baik. Terus pertahankan. Terima kasih	Positif
4	Penghargaan untuk seluruh aspek disini, fasilitas ok dan staffnya ramah. Pilihannya pertahankan atau tingkatkan.	Positif
5	Pokok e de best sekali. Pelayanan ramah serta cepat.... makasih bpjs kesehatan telah membantu saya....	Positif
TOTAL		500

Dari total 500 data terdapat 97,4% yang memiliki sentiment positif, sedangkan 2,6% lainnya memiliki sentiment negatif. Visualisasi distribusi data pada Gambar 3.1.



**Gambar 3.1** Visualisasi Sentimen

### 3.3 Pre-Processing

Sebelum data *review* digunakan, tahap *pre-processing* dilakukan untuk mendapatkan data bersih. Tahapan yang dilakukan antara lain *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. Berikut ini adalah sebelum dan setelah dilakukan *pre-processing*.

#### 3.3.1 Case Folding

Tahap pertama dalam preprocessing adalah *case folding*. *Case folding* adalah proses mengonversi teks menjadi huruf kecil (*lowercase*) untuk menghilangkan variasi antara huruf besar dan huruf kecil dalam analisis teks. Hasil dari data *case folding* pada Tabel 3.3.

**Tabel 3.3** Hasil Data *Case Folding*

No.	<i>Case Folding</i>
1	pelayanan luarbiasa, cepat, tepat.. ramah
2	saya senang dengan pelayanannya bagus dan petugasnya rama dalam menjelaskannya.
3	alhamdulillah pelayanan bagus. tempatnya dingin dan mudah ditemukan. cs, security dan yg bertugas yg lain sangat ramah. sopan dan menjelaskan dgn baik. terus pertahankan. terima kasih
4	penghargaan untuk seluruh aspek disini, fasilitas ok dan staffnya ramah. pilihannya pertahankan atau tingkatkan.
5	pokok e de best sekali. pelayanan ramah serta cepat.... makasih bpjs kesehatan telah membantu saya....
Total	500

### 3.3.2 *Cleaning*

Tahap selanjutnya adalah penghapusan karakter yang tidak diperlukan dari teks. Langkah ini bertujuan untuk membersihkan teks dari karakter-karakter yang tidak relevan dan dapat mengganggu proses analisis.. Hal ini karena ulasan dalam bahasa Indonesia seringkali mengandung simbol, kalimat atau kata yang tidak baku, angka, hashtag, dan tautan URL. Hasil dari data *cleaning* pada Tabel 3.4

**Tabel 3.4** Hasil Data *Cleaning*

No.	<i>Cleaning</i>
1	pelayanan luarbiasa cepat tepat ramah
2	saya senang dengan pelayanannya bagus dan petugasnya rama dalam menjelaskannya
3	alhamdulillah pelayanan bagus tempatnya dingin dan mudah ditemukan cs security dan yg bertugas yg lain sangat ramah sopan dan menjelaskan dgn baik terus pertahankan terima kasih
4	penghargaan untuk seluruh aspek disini fasilitas ok dan staffnya ramah pilihannya pertahankan atau tingkatkan
5	pokok de best sekali pelayanan ramah serta cepat makasih bpjs kesehatan telah membantu saya
Total	500

### 3.3.3 Tokenizing

Tahap selanjutnya dalam *preprocessing* adalah *tokenizing*. Proses ini melibatkan pemecahan kalimat menjadi kata-kata individual atau token. Tujuannya adalah untuk mempersiapkan teks agar dapat dianalisis pada tingkat kata. Hasil dari *tokenizing* pada Tabel 3.5

**Tabel 3.5** Hasil Data *Tokenizing*

No.	<i>Tokenizing</i>
1	['pelayanan', 'luarbiasa', 'cepat', 'tepat', 'ramah']
2	['saya', 'senang', 'dengan', 'pelayanannya', 'bagus', 'dan', 'petugasnya', 'rama', 'dalam', 'menjelaskannya']
3	['alhamdulillah', 'pelayanan', 'bagus', 'tempatnya', 'dingin', 'dan', 'mudah', 'ditemukan', 'cs', 'security', 'dan', 'yg', 'bertugas', 'yg', 'lain', 'sangat', 'ramah', 'sopan', 'dan', 'menjelaskan', 'dgn', 'baik', 'terus', 'pertahankan', 'terima', 'kasih']
4	['penghargaan', 'untuk', 'seluruh', 'aspek', 'disini', 'fasilitas', 'ok', 'dan', 'staffnya', 'ramah', 'pilihannya', 'pertahankan', 'atau', 'tingkatkan']
5	['pokok', 'de', 'best', 'sekali', 'pelayanan', 'ramah', 'serta', 'cepat', 'makasih', 'bpjs', 'kesehatan', 'telah', 'membantu', 'saya']
Total	500

### 3.3.4 Stopword Removal

*Stopword removal* digunakan untuk menghapus kata-kata yang tidak memiliki makna penting dalam teks, sehingga hanya kata-kata yang relevan dan signifikan yang tersisa untuk dianalisis. Hasil *stopword removal* pada Tabel 3.6

**Tabel 3.6** Hasil Data *Stopword Removal*

No.	<i>Stopword Removal</i>
1	['pelayanan', 'luarbiasa', 'cepat', 'rama']
2	['senang', 'pelayanannya', 'bagus', 'petugasnya', 'rama', 'menjelaskannya']
3	['alhamdulillah', 'pelayanan', 'bagus', 'tempatnya', 'dingin', 'mudah', 'ditemukan', 'cs', 'security', 'bertugas', 'ramah', 'sopan', 'dgn', 'pertahankan', 'terima', 'kasih']
4	['penghargaan', 'aspek', 'fasilitas', 'ok', 'staffnya', 'ramah', 'pilihannya', 'pertahankan', 'tingkatkan']
5	['pokok', 'best', 'pelayanan', 'ramah', 'cepat', 'makasih', 'bpjs', 'kesehatan', 'membantu']
Total	500

### 3.3.5 *Stemming*

Tahap *stemming* merupakan proses mengubah kata berimbuhan menjadi akar kata. Pada tahap ini, akar kata dari setiap kata dicari menggunakan modul Sastrawi pada Python. Hasil dari *stemming* dapat dilihat pada Tabel 3.7.

**Tabel 3.7** Hasil Data *Stemming*

No.	<i>Stemming</i> Data
1	layan luarbiasa cepat ramah
2	senang layan bagus tugas rama jelas
3	alhamdulillah layan bagus tempat dingin mudah temu cs security tugas ramah sopan dgn tahan terima kasih
4	harga aspek fasilitas ok staffnya ramah pilih tahan tingkat
5	pokok best layan ramah cepat makasih bpjs sehat bantu
Total	500

### 3.4 Pembobotan Kata (TF-IDF)

Pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan untuk mengidentifikasi kata-kata penting dalam teks ulasan BPJS Kesehatan Samarinda di Google Maps. Dataset yang digunakan terdiri dari 500 baris dan 463 kolom, di mana setiap kolom mewakili kata unik dalam korpus data. Berdasarkan hasil Gambar 3.2 TF-IDF, terlihat bahwa sebagian besar kata memiliki bobot rendah senilai 0, yang menunjukkan bahwa kata-kata tersebut jarang muncul dalam dokumen.

```

adandiskriminasi admin administrasi adminitrasi adu ahamdulillah \
0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
1 0.0 0.0 0.0 0.0 0.0 0.0 0.0
2 0.0 0.0 0.0 0.0 0.0 0.0 0.0
3 0.0 0.0 0.0 0.0 0.0 0.0 0.0
4 0.0 0.0 0.0 0.0 0.0 0.0 0.0
.. .. .. .. .. .. ..
495 0.0 0.0 0.0 0.0 0.0 0.0 0.0
496 0.0 0.0 0.0 0.0 0.0 0.0 0.0
497 0.0 0.0 0.0 0.0 0.0 0.0 0.0
498 0.0 0.0 0.0 0.0 0.0 0.0 0.0
499 0.0 0.0 0.0 0.0 0.0 0.0 0.0

aj aja akses aktif ... via warga wassalamu wb wr yaa yna \
0 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
1 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
2 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
3 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
4 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
.. .. .. .. .. .. ..
495 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
496 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
497 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
498 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
499 0.0 0.0 0.0 0.0 ... 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

you yuk sentimen
0 0.0 0.0 positif
1 0.0 0.0 positif
2 0.0 0.0 positif
3 0.0 0.0 positif
4 0.0 0.0 positif
.. .. .. .. .. .. ..
495 0.0 0.0 positif
496 0.0 0.0 positif
497 0.0 0.0 positif
498 0.0 0.0 positif
499 0.0 0.0 positif

[500 rows x 463 columns]

```

**Gambar 3.2** Pembobotan Kata (TF-IDF)

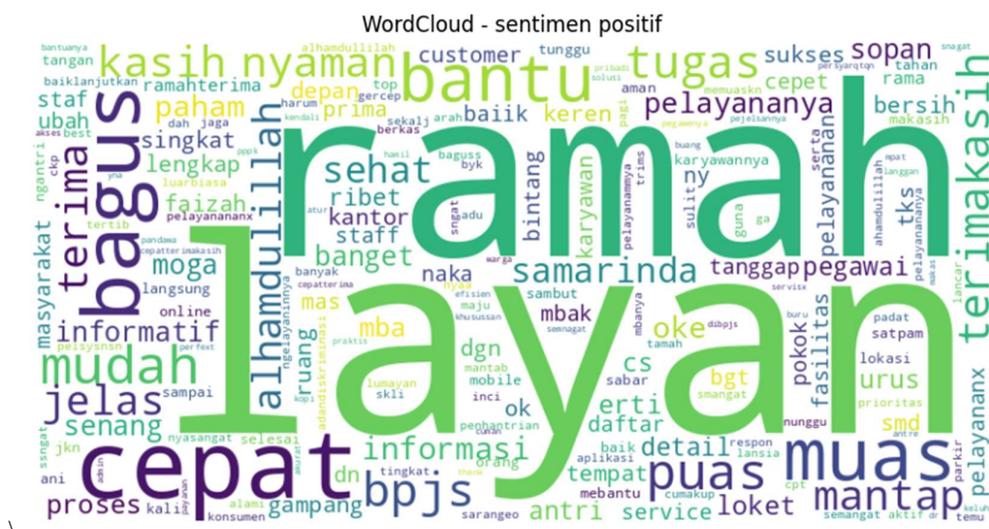
Berdasarkan hasil analisis *Term Frequency* (TF), ditemukan kata-kata yang paling sering muncul dan memiliki bobot signifikan dalam korpus data. Gambar 3.3 menunjukkan bahwa kata "layan" memiliki frekuensi tertinggi dengan skor 118.940490, diikuti oleh kata "ramah" dengan skor 88.053426, dan "cepat" dengan skor 60.275206. Kata-kata lain yang juga memiliki bobot tinggi adalah "bagus" dengan skor 40.736076, "muas" 33.703185, "bantu" 26.888281, "mudah" 17.03212, "puas" 15.745396, "bpjs" 14.942416, dan "tugas" 12.741422. Frekuensi tinggi dari kata-kata ini menunjukkan topik yang dominan dan relevan dalam ulasan BPJS Kesehatan Samarinda di Google Maps.

Most frequent terms:		
	term	frequency
205	layan	118.940490
329	ramah	88.053426
74	cepat	60.275206
32	bagus	40.736076
251	muas	33.703185
40	bantu	26.888281
252	mudah	17.032124
323	puas	15.745396
59	bpjs	14.942416
437	tugas	12.741422

Gambar 3.3 Scoring TF

### 3.5 Visualisasi

Kata-kata yang sering muncul dalam teks ulasan divisualisasikan menggunakan *wordcloud*. Dalam pembuatan *wordcloud*, peneliti menggunakan *library wordcloud* dan hasil ekstraksi fitur dengan metode TF-IDF. Semakin sering kata tersebut muncul dalam data, semakin besar ukurannya dalam *wordcloud*. Visualisasi kata-kata yang sering muncul dalam ulasan positif pada Gambar 3.4, sedangkan visualisasi kata-kata yang sering muncul dalam ulasan negatif pada Gambar 3.5.



Gambar 3.4 WordCloud Positif



untuk pengujian dengan rasio 80:20, dan 99.3% untuk pengujian dengan rasio 70:30. Pengujian ini menunjukkan bahwa model *K-Nearest Neighbor* yang dilatih dengan teknik *preprocessing* teks yang tepat serta pembobotan TF-IDF memberikan performa yang cukup baik. Hasil evaluasi menunjukkan bahwa model *K-Nearest Neighbor* dengan rasio pengujian 70:30 mencapai akurasi sebesar 0.993, yang berarti model berhasil mengklasifikasikan sekitar 99.3% data dengan benar. Pada rasio pengujian 80:20, model menunjukkan penurunan performa dengan akurasi sebesar 0.99, mengindikasikan bahwa sekitar 99% data diklasifikasikan dengan benar. Sementara pada rasio pengujian 90:10, akurasi model menunjukkan semakin menurun menjadi 0.98, yang berarti model mampu mengklasifikasikan sekitar 98% data dengan benar.

Perbedaan akurasi ini menunjukkan bahwa performa model sedikit bervariasi tergantung pada rasio pembagian data latih dan data uji. Pada rasio 70:30 dan 80:20, model menunjukkan performa yang lebih baik dibandingkan dengan rasio 90:10. Hal ini dapat disebabkan oleh berbagai faktor, seperti jumlah data latih yang lebih besar pada rasio 70:30 dan 80:20, yang memungkinkan model untuk mempelajari lebih banyak pola dari data. dan pada rasio 90:10, jumlah data latih yang lebih sedikit mungkin tidak cukup untuk menangkap variasi yang ada dalam data, sehingga menyebabkan penurunan akurasi. Secara keseluruhan, hasil ini menunjukkan pemilihan rasio yang tepat dalam pembagian data latih dan data uji untuk mendapatkan performa model yang optimal. Pemilihan rasio yang baik dapat membantu model mempelajari pola dengan lebih baik dan menghasilkan prediksi yang lebih akurat (Kurniawan & Anubhakti, 2023). Dan dari ketiga rasio tersebut peneliti mengambil rasio yang terbaik yaitu 70:30 dimana 70% sebagai data latih sebanyak 350 data dan 30% data uji sebanyak 150 data.

Setelah memilih rasio terbaik kemudian dilakukan pengujian untuk mencari akurasi terbaik pada nilai *K* dalam algoritma *K-Nearest Neighbors* (KNN), perhitungan jarak *K* yang digunakan menggunakan *Euclidean Distance*. Penggunaan beberapa nilai *K* bertujuan untuk membandingkan hasil akurasi dari setiap nilai *K*. Hal ini dilakukan baik saat uji validasi

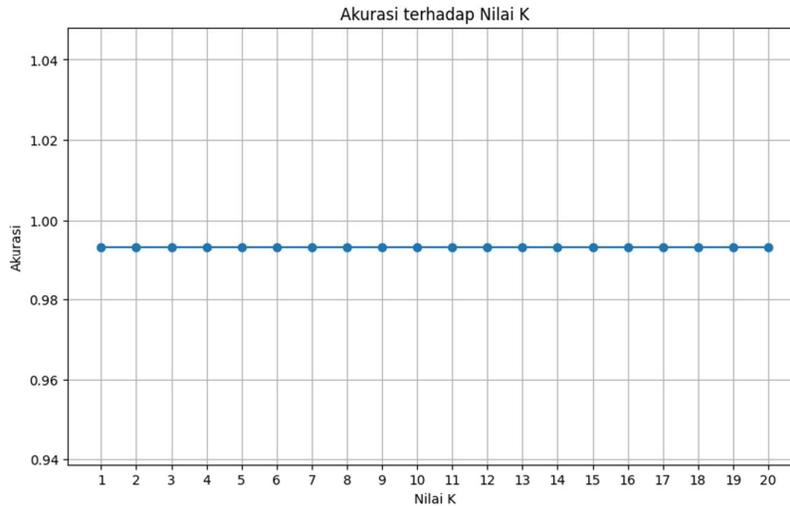
maupun evaluasi akhir untuk menentukan nilai K yang paling optimal di antara semua nilai yang diuji. Dengan membandingkan akurasi pada setiap nilai K, kita dapat mengetahui nilai K yang memberikan kinerja terbaik untuk model tersebut. (Herman et al., 2020)

Hasil evaluasi akurasi dari model *K-Nearest Neighbor* (KNN) dengan berbagai nilai K dari 1 hingga 20. Setiap nilai K menghasilkan akurasi yang sama, yaitu 0.9933. Menunjukkan bahwa model KNN memiliki performa yang sangat baik dan stabil pada dataset yang digunakan, dan perubahan nilai K tidak mempengaruhi akurasi secara signifikan karena adanya ketidakseimbangan pada kelas di mana salah satu kelas mendapatkan jumlah sampel yang lebih banyak dibandingkan dengan kelas lainnya. Ketidakseimbangan ini bisa menyebabkan model lebih mudah untuk memprediksi kelas mayoritas dengan benar, sehingga meningkatkan akurasi keseluruhan tanpa benar-benar mencerminkan kemampuan model dalam memprediksi kelas minoritas dengan akurat (Nur Ariyanti & Cahya Wihandika, 2019). Hasilnya pada Gambar 3.6.

```
Akurasi untuk berbagai nilai K:  
Nilai K: 1, Akurasi: 0.9933  
Nilai K: 2, Akurasi: 0.9933  
Nilai K: 3, Akurasi: 0.9933  
Nilai K: 4, Akurasi: 0.9933  
Nilai K: 5, Akurasi: 0.9933  
Nilai K: 6, Akurasi: 0.9933  
Nilai K: 7, Akurasi: 0.9933  
Nilai K: 8, Akurasi: 0.9933  
Nilai K: 9, Akurasi: 0.9933  
Nilai K: 10, Akurasi: 0.9933  
Nilai K: 11, Akurasi: 0.9933  
Nilai K: 12, Akurasi: 0.9933  
Nilai K: 13, Akurasi: 0.9933  
Nilai K: 14, Akurasi: 0.9933  
Nilai K: 15, Akurasi: 0.9933  
Nilai K: 16, Akurasi: 0.9933  
Nilai K: 17, Akurasi: 0.9933  
Nilai K: 18, Akurasi: 0.9933  
Nilai K: 19, Akurasi: 0.9933  
Nilai K: 20, Akurasi: 0.9933
```

**Gambar 3.6** Hasil Akurasi Nilai K

Plot yang menunjukkan hubungan antara nilai K dan akurasi dalam model *K-Nearest Neighbor* (KNN). Sumbu horizontal merepresentasikan nilai K, yang berkisar dari 1 hingga 20. Sumbu vertikal merepresentasikan akurasi model. Dari plot tersebut, terlihat bahwa akurasi tetap sama untuk setiap nilai K yang diuji. Hasilnya pada Gambar 3.7.



**Gambar 3.7** Plot Akurasi Nilai K

Hasil evaluasi akurasi dari model K-Nearest Neighbor (KNN) dengan berbagai nilai K, dari 1 hingga 20. Setiap nilai K menghasilkan akurasi yang sama, yaitu 0.9933. Menunjukkan bahwa model KNN memiliki performa yang sangat baik dan stabil pada dataset yang digunakan, dan perubahan nilai K tidak mempengaruhi akurasi secara signifikan. Stabilitas akurasi ini mengindikasikan bahwa model KNN sangat efektif dalam memprediksi label data uji dengan tingkat ketepatan yang konsisten. Meskipun hasil akurasi yang diperoleh sama untuk semua nilai K yang diuji, peneliti memutuskan untuk menggunakan nilai K=5 sebagai parameter final untuk model KNN.

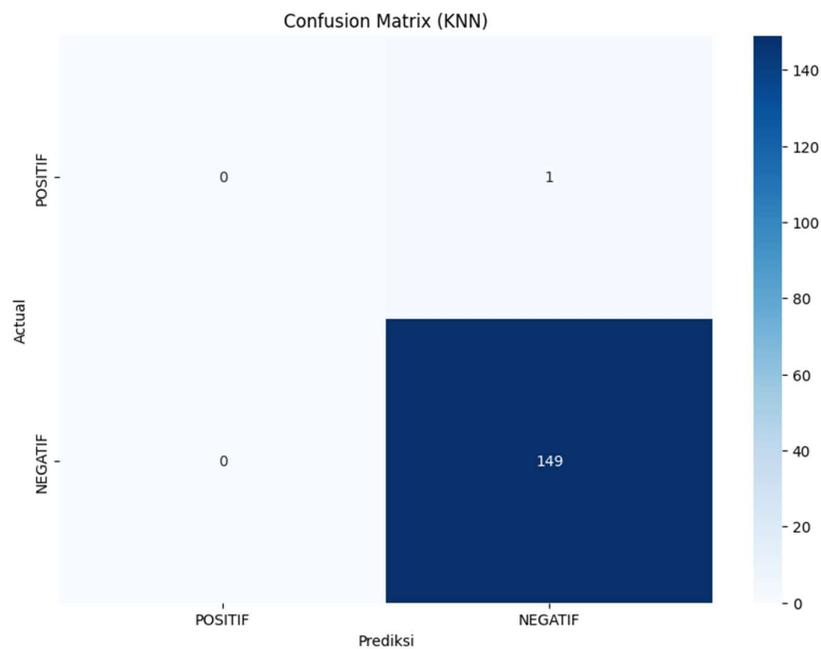
Untuk mengevaluasi kinerja suatu model klasifikasi yang memprediksi sentimen dari teks. Evaluasi ini melibatkan perhitungan nilai-nilai seperti akurasi (*Accuracy*), Empat variabel (TN, TP, FN, FP) mewakili nilai dari setiap kategori dalam *confusion matrix*, yaitu *True Negative* (TN), *True Positive* (TP), *False Negative* (FN) dan *False Positive* (FP). Kemudian, dilakukan perhitungan persentase dari masing-masing kategori tersebut terhadap total jumlah data. Visualisasi yang dihasilkan merupakan nilai *confusion matrix* dengan label masing-masing kategori dan peneliti mengambil rasio yang terbaik yaitu 70:30 dimana 70% sebagai

data latih sebanyak 350 data dan 30% data uji sebanyak 150 data dan nilai  $K=5$ . Akurasi model dihitung berdasarkan prediksi yang dihasilkan pada data uji pada Gambar 3.8.

**Gambar 3.8** Hasil Akurasi Metode KNN

Accuracy: 0.9933333333333333

Akurasi dari model diatas adalah akurasi yang digunakan oleh peneliti menggunakan data yang diberi label oleh (*expert*) sebesar 0.993 atau 99.3%. Untuk tampilan *Confusion Matrix* akan ditampilkan pada Gambar 3.9.



**Gambar 3.9** Hasil *Confusion Matrix*

Adapun perhitungan manual berdasarkan hasil Gambar 3.9 pada Tabel 3.9.

**Tabel 3.9** Hasil Data Uji

		True Postive	False Negatif
Prediction	Postive	0	1
	Negatif	0	149
Total		150	

Akurasi adalah rasio dari jumlah prediksi yang benar (*True Positif dan True Negatif*) terhadap total jumlah prediksi.

Berdasarkan *Confusion Matriks* di atas :

- a) *True Positif* (TP) = 0
- b) *True Negatif* (TN) = 149
- c) *False Positif* (FP) = 1
- d) *False Negatif* (FN) = 0

Hasil perhitungan manual akurasi pada Persamaan (3.1).

$$Akurasi = \frac{0+149}{0+149+1+0} = \frac{149}{150} = 0.993\% \quad (3.1)$$

Hasil pengujian model menunjukkan akurasi sebesar 0,993 atau 99.3% pada data uji. Berarti dari 150 data yang diuji, model mampu memprediksi dengan benar 99 kasus. Akurasi ini dihitung berdasarkan jumlah prediksi yang benar (*True Positive dan True Negative*) dibagi dengan total jumlah kasus.

## **BAB IV**

### **PENUTUP**

#### **4.1 Simpulan**

Berdasarkan hasil penelitian analisis sentimen pada ulasan Google Maps mengenai layanan BPJS Kesehatan di Samarinda yang menggunakan metode *K-Nearest Neighbor* dengan ekstraksi fitur TF-IDF, dapat disimpulkan bahwa mayoritas ulasan memiliki sentimen positif, yaitu sebesar 97,4%, sementara ulasan dengan sentimen negatif hanya 2,6%, *labelling data* dilakukan oleh seorang ahli bahasa (*expert*). Tahapan pre-processing seperti case folding, cleaning, tokenizing, stopword removal, dan stemming terbukti penting dalam mempersiapkan data ulasan untuk analisis lebih lanjut. Pembobotan TF-IDF dalam ekstraksi fitur terbukti efektif dalam mengidentifikasi kata-kata penting dan relevan dari ulasan pengguna. Teknik ini membantu menangkap informasi yang paling penting dan berkontribusi signifikan terhadap sentimen yang diekspresikan dalam ulasan, dengan kata-kata yang sering muncul memiliki bobot tinggi terkait pengalaman dan layanan BPJS Kesehatan di Samarinda.

Model *K-Nearest Neighbor* yang diterapkan, dengan rasio data latih dan data uji sebesar 70:30 dan nilai  $k=5$ , menunjukkan tingkat akurasi yang sangat tinggi, yaitu 99,3%. Hasil ini menunjukkan bahwa metode *K-Nearest Neighbor* mampu mengklasifikasikan opini pengguna dengan sangat baik, membuktikan kemampuannya dalam memprediksi sentimen dengan tingkat kesalahan yang rendah. Hasil penelitian ini diharapkan dapat digunakan sebagai dasar untuk meningkatkan kualitas layanan BPJS Kesehatan di Samarinda melalui pemahaman yang lebih baik terhadap umpan balik pengguna. Penelitian ini juga menunjukkan potensi penerapan teknik serupa untuk analisis sentimen pada berbagai layanan publik lainnya.

#### **4.2 Implikasi**

Terdapat beberapa implikasi dari penelitian ini, Pertama, bagi BPJS Kesehatan Samarinda, hasil analisis sentimen ini dapat digunakan sebagai umpan balik yang berharga untuk

mengevaluasi dan meningkatkan kualitas pelayanan mereka. Identifikasi aspek-aspek yang sering mendapat ulasan negatif memungkinkan BPJS untuk fokus pada perbaikan di area tersebut, sehingga dapat meningkatkan kepuasan peserta secara keseluruhan. Kedua, bagi masyarakat, penelitian ini menyediakan informasi yang transparan mengenai persepsi publik terhadap layanan BPJS Kesehatan. Terakhir, bagi peneliti, penelitian ini memberikan kontribusi dalam penelitian selanjutnya dengan menunjukkan efektivitas metode *K-Nearest Neighbor* dalam klasifikasi sentimen menggunakan fitur ekstraksi TF-IDF. Hal ini membuka peluang untuk penelitian lanjutan yang dapat memperdalam pemahaman tentang analisis sentimen dan pengembangan metode yang lebih akurat dalam konteks yang sama.

## DAFTAR RUJUKAN

- Al Khadafi, M., Paranita Kartika, K., & Febrinita, F. (2022). PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN LEXICON BASED UNTUK ANALISIS SENTIMEN CYBERBULLYING PADA BPJS. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 2).
- Amanny Ulfah Nabiyah Ramadhanty, I. N. (2023). *IMPLEMENTASI WEB SCRAPING PADA SITUS JURNAL SINTA MENGGUNAKAN FRAMEWORK SELENIUM WEBDRIVER PYTHON*.
- Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1). <https://doi.org/10.30812/bite.v2i1.804>
- Annisa Medina Sari. (2023, December 14). *Perbedaan Fakes BPJS Tingkat 1,2,3 Yang Harus Diketahui*.
- Antara Kaltim. (2023, March 17). *Tingkat kepesertaan BPJS Kesehatan Samarinda mencapai 100,89 persen*.
- Assidyk, A. N., Setiawan, E. B., Si, S., Kurniawan, I., Pd, S., & Si, M. (2020). *Analisis Perbandingan Pembobotan TF-IDF dan TF-RF pada Trending Topic di Twitter dengan Menggunakan Klasifikasi K-Nearest Neighbor*.
- Azzahra Nasution, D., Khotimah, H. H., & Chamidah, N. (2019). *PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN* (Vol. 4, Issue 1).
- BPJS Kesehatan. (n.d.). <https://bpjs-kesehatan.go.id/>.
- CNN INDONESIA. (2021, November 9). *Kisah Pengidap Tumor Asal Sules Lancar Operasi Berkat JNN-KIS*.
- Fadiyah Basar, T., Ratnawati, D. E., & Arwani, I. (2022). *Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest* (Vol. 6, Issue 3). <http://j-ptiik.ub.ac.id>
- Herman, I. H., Widiyanto, D., & Ernawati, I. (2020). *PENGUNAAN K-NEAREST NEIGHBOR (KNN) UNTUK MENGIDENTIFIKASI CITRA BATIK PEWARNA ALAMI DAN PEWARNA SINTETIS BERDASARKAN WARNA*.
- IDN Times Kaltim. (2019, August 19). *Ribet, Warga Keluhkan Pelayanan BPJS Kesehatan di Samarinda*.
- Khadijah Nur Azizah. (2021, March 9). *7 Tahun BPJS Kesehatan, Peserta Masih Keluhkan Antrean Berjam-jam*.
- Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 406. <https://doi.org/10.30865/mib.v5i2.2835>
- Kurniawan, R., & Anubhakti, D. (2023). *IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINEDALAM MEMREDIKSI HARGA SAHAM PT. KRAKATAU STEEL TBK*. 2(2).
- Linda, L., Haskas, Y., Kadrianti, E., & Nani Hasanuddin Makassar, S. (n.d.). *PERBEDAAN PERSEPSI PENGGUNA JASA BPJS DAN NON BPJS (UMUM) TENTANG*

- KUALITAS PELAYANAN KEPERAWATAN DIRSUD TIMIKA-PAPUA. In *Jurnal Ilmiah Kesehatan Diagnosis* (Vol. 15).
- Nur Ariyanti, R., & Cahya Wihandika, R. (2019). *Identifikasi Jenis Attention Deficit Hyperactivity Disorder (ADHD) Pada Anak Usia Dini Menggunakan Metode Modified K-Nearest Neighbor (MKNN)* (Vol. 3, Issue 1). <http://j-ptiik.ub.ac.id>
- Parasian Doloksaribu, H., & Samuel, Y. T. (2022). *KOMPARASI ALGORITMA DATA MINING UNTUK ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI*. 16(1). <https://doi.org/10.47111/JTI>
- Puspita, R., & Widodo, A. (2021a). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- Puspita, R., & Widodo, A. (2021b). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- Putra, P., H Pardede, A. M., & Syahputra, S. (2022). ANALISIS METODE K-NEAREST NEIGHBOUR (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA. *Jurnal Teknik Informatika Kaputama (JTik)*, 6(1).
- Putri, A., Syaficha Hardiana, C., Novfuja, E., Try Puspa Siregar, F., Fatma, Y., & Wahyuni, R. (2023). Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir. *Institut Riset Dan Publikasi Indonesia (IRPI) MALCOM: Indonesian Journal of Machine Learning and Computer Science Journal Homepage*, 3(1), 20–26.
- Ratih Puspitasari, Findawati, Y., & Rosid, M. A. (2023). SENTIMENT ANALYSIS OF POST-COVID-19 INFLATION BASED ON TWITTER USING THE K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE CLASSIFICATION METHODS. *Jurnal Teknik Informatika (Jutif)*, 4(4), 669–679. <https://doi.org/10.52436/1.jutif.2023.4.4.801>
- Rifa, M. K., Totohendarto, M. H., & Muttaqin, M. R. (2023). Analisis Sentimen Pengguna E-Wallet Dana Dan Gopay Pada Twitter Menggunakan Metode Support Vector Machine (SVM). *IJCCS*, x, No.x, 1–5.
- Salim, S. S., & Mayary, J. (2020). ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP DOMPET ELEKTRONIK DENGAN METODE LEXICON BASED DAN K – NEAREST NEIGHBOR. *Jurnal Ilmiah Informatika Komputer*, 25(1), 1–17. <https://doi.org/10.35760/ik.2020.v25i1.2411>

# LAMPIRAN

## Lampiran 1 CV *Expert Labeling*

### **Irfan Abdul Hakim**

Hirfan825@gmail.com/082135357602

#### **PROFIL**

Saya adalah seorang profesional yang memiliki keterampilan komunikasi yang baik, pemikiran terbuka, dan selalu bersemangat dalam mencari pengetahuan baru. Saya memiliki keinginan yang besar untuk terus belajar dan berkembang. Saya dengan cepat dapat menyesuaikan diri dengan perubahan dan memiliki kemampuan untuk beradaptasi dengan cepat dalam lingkungan kerja yang dinamis. Saya juga sangat berpengetahuan dalam penggunaan teknologi dan memiliki kemampuan untuk memanfaatkannya secara efektif. Saya mencari peluang untuk bekerja dalam lingkungan yang inovatif, di mana saya dapat menerapkan keterampilan analitis dan berpikir kritis saya untuk mencapai hasil yang optimal.

#### **PENDIDIKAN**

**UNIVERSITAS GADJAH MADA**

2015-2020

**S1 Sosiologi, IPK: 3.40/4.00**

**S2 Sosiologi, IPK: 3.50/ 4.00**

#### **PENGALAMAN KERJA**

**SMA NURUL MUSLIM BATEALIT JEPARA**

**STAF PENGAJAR SOSIOLOGI**

- Merancang Silabus pembelajaran di Awal Semester.
- Melakukan aktifitas belajar mengajar sesuai dengan rancangan awal silabus.
- Melakukan evaluasi hasil pembelajaran melalui penilaian harian, penilaian tengah semester dan penilaian akhir semester.

**PANITIA PENGAWAS PEMILU (PANWASLU) KECAMATAN SEMBORO**

2022 - sekarang

**STAF PELAKSANA**

- Membantu Tugas Pengawasan komisioner panwaslu kecamatan.
- Melakukan pendataan laporan dan temuan pelanggaran baik dari aduan masyarakat maupun temuan langsung Panwaslu.
- Melakukan Pengarsipan Surat Keluar Masuk Laporan dan Temuan Pelanggaran.

**MTS ALI MAKSUM YOGYAKARTA**

2021 - 2022

**STAF PENGAJAR**

- Merancang Silabus pembelajaran di Awal Semester.
- Melakukan aktifitas belajar mengajar sesuai dengan rancangan awal silabus.
- Melakukan evaluasi hasil pembelajaran melalui penilaian harian, penilaian tengah semester dan penilaian akhir semester.

#### **PENGALAMAN ORGANISASI**

**GERAKAN MAHASISWA SATU BANGSA (GEMASABA) KAB. SLEMAN**

2021 - 2022

**WAKIL KETUA 1**

- Membantu Ketua dalam Pembuatan, pelaksanaan serta pengawasan terhadap berjalannya program kerja.

- Menjalin hubungan baik dengan organisasi lain.
- Melakukan koordinasi dan inisiasi kerja sama dengan organisasi lain untuk kemajuan organisasi.

**PENGURUS ASRAMA MTS PUTRA PONDOK PESANTREN KRAPYAK YAYASAN ALI MAKSUM 2015 - 2022**

**KOORDINATOR BENDAHARA**

- Merancang, mengelola dan mendistribusikan Anggaran tahunan untuk tiga asrama
- Melakukan koordinasi dengan seluruh divisi terkait kebutuhan anggaran dan pendistribusiannya.
- Melakukan koordinasi langsung dengan bendahara yayasan terkait dengan anggaran asrama.
- Melakukan Pelaporan rutin terhadap penggunaan anggaran kepada pengasuh pondok pesantren.

**IKATAN ALUMNI MA ALI MAKSUM YOGYAKARTA ANGKATAN 2015**

2018 - sekarang

**KETUA**

- Merancang serta merealisasikan program kerja tahunan.
- Menjaga hubungan baik dengan seluruh alumni.
- Melakukan koordinasi terkait agenda - agenda terdekat.

**PERGERAKAN MAHASISWA ISLAM INDONESIA (PMII) GADJAH MADA**

2018 - 2019

**KOORDINATOR DIVISI KADERISASI**

- Membuat dan menjalankan bersama tim program kerja yang telah disusun.
- Melakukan koordinasi dengan pengurus rayon divisi kaderisasi.
- Merancang strategi rekrutmen yang efektif dan tepat sasaran di awal tahun ajaran baru.

**PENGALAMAN KESUKARELAAN**

- **Survey Kepuasan Pelanggan PDAM Yogyakarta** - Yogyakarta ( Februari 2022): Koordinator Enumerator
- **Seminar Nasional Menjadi Wirausahawan Sukses** - Yogyakarta ( September 2020): Moderator
- **Penerimaan Santri Baru Yayasan Ali Maksu** - Yogyakarta ( Januari - Februari 2020): Divisi Logistik
- **Riset Pusat Studi Pancasila Universitas Pembangunan Nasional Yogyakarta** - Yogyakarta ( Agustus-September 2019): Enumerator
- **Pelatihan Kader Dasar PMII Gajah Mada** - (Maret 2018): Sterring Comitte
- **Pekan Raya Anak** - Yogyakarta ( Desember 2017): Bendahara
- **Kulo Nuwun Party** - Yogyakarta ( Maret 2016): Divisi Logistik

**PENGALAMAN RISET**

- Penulisan Buku Jusuf Kalla: Sang Pendamai Ulung
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Kopi Sianida
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci IKN
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Bitcoin Halving
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Hak Angket
- Anotasi Data Sentiment Tweet [Analisis Sentiment Neural Network] dengan Kata Kunci Quick Count

## Lampiran 2 Surat Izin Penelitian

 UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR Berkeadilan   Berprestasi   Berkeagamaan	<b>UMKT</b> Program Studi Teknik Informatika Fakultas Sains dan Teknologi	Telp. 0541-748511 Fax. 0541-766832 Website <a href="http://informatika.umkt.ac.id">http://informatika.umkt.ac.id</a> email: <a href="mailto:informatika@umkt.ac.id">informatika@umkt.ac.id</a>
---	--	--

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Nomor : 055-005/KET/FST.1/A/2024  
Lampiran : -  
Perihal : Keterangan Pengambilan Data Sekunder

*Assalamu'alaikum Warrahmatullahi Wabarrakatuh*

Puji Syukur kepada Allah Subhanahu wa ta'ala yang senantiasa melimpahkan Rahmat-Nya kepada kita sekalian. Amin.

Dengan surat ini, kami menerangkan bahwa mahasiswa berikut:

Nama : Ikhsan Nuttakwa Takbirata Ihram Nabawi  
NIM : 2011102441072  
Program Studi : Teknik Informatika

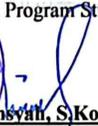
Melakukan penelitian dengan pengambilan data sekunder di Google Maps data yang diambil yaitu data ulasan pelayanan BPJS.

Demikian hal ini disampaikan, atas kerjasamanya kami ucapkan terima kasih.

*Wassalamu'alaikum Warrahmatullahi Wabarrakatuh*

Samarinda, 20 Dzulhijjah 1445 H  
27 Juni 2024 M

Ketua Program Studi S1 Teknik Informatika

  
**Arhansyah, S.Kom., M.TI**  
IDN. 1118019203



Kampus 1 : Jl. Ir. H. Juanda, No.15, Samarinda  
Kampus 2 : Jl. Pelita, Pesona Mahakam, Samarinda

CS Dipindai dengan CamScanner

### Lampiran 3 Pengumpulan Data

```
[1] feature_id = "0x2df67f574f743fe1:0x9d1617309fd1d93a"
    loop = 50

pip install beautifulsoup4
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (4.12.3)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4) (2.5)

[3] from datetime import datetime, timedelta

# nama hari dalam bahasa Indonesia
nama_hari_indonesia = {
    "Monday": "Senin",
    "Tuesday": "Selasa",
    "Wednesday": "Rabu",
    "Thursday": "Kamis",
    "Friday": "Jumat",
    "Saturday": "Sabtu",
    "Sunday": "Minggu"
}
```

```
# Fungsi untuk mengonversi teks waktu menjadi timedelta
def konversi_teks_ke_timedelta(teks):
    if 'hari' in teks:
        jumlah = int(teks.split()[0])
        return timedelta(days=jumlah)
    elif 'minggu' in teks:
        if 'seminggu' in teks:
            return timedelta(weeks=1)
        else:
            jumlah = int(teks.split()[0])
            return timedelta(weeks=jumlah)
    elif 'bulan' in teks:
        if 'sebulan' in teks:
            return timedelta(days=30)
        else:
            jumlah = int(teks.split()[0])
            return timedelta(days=jumlah*30)
    else:
        raise ValueError("Format teks waktu tidak dikenali")

# Fungsi untuk mengonversi teks waktu menjadi tanggal dan hari
def konversi_teks_ke_tanggal(teks):
    if 'hari' in teks or 'minggu' in teks or 'bulan' in teks:
        timedelta_lalu = konversi_teks_ke_timedelta(teks)
        tanggal = datetime.now() - timedelta_lalu
        nama_hari = tanggal.strftime("%A")
        nama_hari_indo = nama_hari_indonesia.get(nama_hari, nama_hari)
        return tanggal, nama_hari_indo
    elif 'tahun' in teks:
        if 'setahun' in teks:
```

```

        return datetime.now() - timedelta(days=365), None
    else:
        jumlah_tahun = int(teks.split()[0])
        return datetime.now() - timedelta(days=365 * jumlah_tahun), None
    else:
        # Jika format tidak sesuai dengan yang dikenali, kembalikan None
        return None, None

import requests
import re
from datetime import datetime, timedelta
from bs4 import BeautifulSoup
import pandas as pd

def get_reviews_data(feature_id):
    token = ''
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/101.0.4951.54 Safari/537.36"
    }

    base_url = "https://www.google.com/async/reviewDialog?hl=id&async=feature_id:{feature_id},next_page_token:{token},sort by:newest,start index:,associated topic:, fmt:pc"

    # List untuk menyimpan DataFrame dari setiap permintaan
    dfs = []

```

```

# Lakukan loop permintaan
for i in range(loop):
    response = requests.get(base_url.format(feature_id=feature_id, token=token), headers=headers)
    soup = BeautifulSoup(response.content, 'html.parser')

    user = []
    location_info = {}
    data_id = ''

    for el in soup.select('.c9QyIf'):
        data_id = soup.select_one('.DvzRrc')['data-fid']
        next_page_token = soup.select_one('.gws-localreviews__general-reviews-block')['data-next-page-token']
        if next_page_token:
            token = next_page_token
        else:
            token = None # Jika token berikutnya kosong, hentikan loop
            break

    for el in soup.select('.gws-localreviews__google-review'):
        # try:
        rating_label = el.select_one('span.lTi8oc.z3Hnkc')['aria-label']
        match = re.search(r'(\d+)(?:,\d+)?', rating_label)

```

```

if match:
    rating = match.group(1)
else:
    rating = "0"
# except KeyError:
#     rating = "0"

waktu = el.select_one('.dehysf').text.strip()
# print(waktu)
tanggal, nama_hari_indo = konversi_teks_ke_tanggal(waktu)
# print(tanggal)
# if tanggal is None: # Tambahkan penanganan eksepsi di sini
#     continue # Lewati data yang tidak valid

tanggal_formatted = tanggal.strftime("%d-%m-%Y")
time_formatted = f"{nama_hari_indo} {tanggal_formatted}"
sentimen = 'Negatif' if int(rating) <= 3 else 'Positif'
user.append({
    'name': el.select_one('.TSuDb').text.strip(),
    'rating': rating,
    'review': el.select_one('.Jtu6Td').text.strip(),
    'time': time_formatted,
    'sentimen': sentimen
})

```

```

df = pd.DataFrame(user)

# Tambahkan DataFrame ke dalam list
dfs.append(df)
total_data += len(df)

print(f"Loop ke-{i+1}:")
# print(df)
print("-----")

if token is None:
    break # Hentikan loop jika token berikutnya kosong

# Gabungkan semua DataFrame menjadi satu
combined_df = pd.concat(dfs, ignore_index=True)

# Simpan DataFrame gabungan ke dalam file Excel
combined_df.to_excel("review_data.xlsx", index=False)

print("Total data yang sudah didapat:", total_data)
print("Data telah disimpan dalam review_data.xlsx")

get_reviews_data(feature_id)

```

#### Lampiran 4 *Install Library*

```

import re # Modul untuk regular expression
import pandas as pd # Manipulasi data tabular dengan Pandas
import numpy as np # Operasi numerik dengan NumPy
import nltk # Pemrosesan teks dengan NLTK

!pip install Sastrawi # Paket Sastrawi untuk pemrosesan bahasa Indonesia

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory # Stemming bahasa Indonesia dari Sastrawi
from nltk.corpus import stopwords # Stopwords dari NLTK
from nltk.tokenize import word_tokenize # Tokenisasi teks dari NLTK

nltk.download('punkt') # Model tokenisasi teks dari NLTK
nltk.download('stopwords') # Daftar stopwords dari NLTK

Requirement already satisfied: Sastrawi in /usr/local/lib/python3.10/dist-packages (1.0.1)
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
True

```

#### Lampiran 5 Membaca Dataset

```

[41] df = pd.read_csv("Review_Data_Baru.csv", encoding='latin1')
df.head(500)

```

## Lampiran 6 Cek Informasi Dataset

```
[42] import pandas as pd

# Mencoba membaca file CSV dengan encoding 'latin1'
data = pd.read_csv("Review_Data_Baru.csv", encoding='latin1')
data.info() # Menampilkan informasi tentang struktur data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   name        500 non-null    object
 1   rating      500 non-null    int64
 2   review      500 non-null    object
 3   time        500 non-null    object
 4   sentimen    500 non-null    object
dtypes: int64(1), object(4)
memory usage: 19.7+ KB
```

## Lampiran 7 Cek Jumlah Sentimen Positif dan Negatif

```
[43] sentiment_counts = data['sentimen'].value_counts() # Menghitung Jumlah Kemunculan sentimen positif dan negatif

positif = sentiment_counts.get('positif', 0)
negatif = sentiment_counts.get('negatif', 0)

print(f"Total Sentimen Positif: {positif}") # mencetak jumlah total sentimen positif.
print(f"Total Sentimen Negatif: {negatif}") # mencetak jumlah total sentimen negatif.
```

```
Total Sentimen Positif: 487
Total Sentimen Negatif: 13
```

## Lampiran 8 Persentase Sentimen

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Misalkan dataset Anda bernama 'Review_Data_Baru' dan kolom targetnya bernama 'sentimen'
df = pd.read_csv('Review_Data_Baru.csv', encoding='latin1')

# Menghitung jumlah sentimen
sentimen_counts = df['sentimen'].value_counts()

# Membuat pie chart untuk persentase sentimen
plt.figure(figsize=(10, 8))
colors = sns.color_palette('viridis', len(sentimen_counts))
sentimen_counts.plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=colors, labels=['Positif', 'Negatif'], wedgeprops={'edgecolor': 'black'})

plt.title('Persentase Sentimen', fontsize=16)
plt.ylabel('') # Menghilangkan label default pada y-axis
plt.legend(title='sentimen', loc='upper left')
plt.show()
```

## Lampiran 9 Membaca Dataset Kolom Review dan Sentimen

```
[45] columns = ['review', 'sentimen']
df = pd.DataFrame(data, columns=columns).iloc[0:500] # membatasi DataFrame hanya pada 500 baris pertama.
df.head(500) # Menampilkan 5 baris pertama dari DataFrame df untuk melihat preview data.
```

## Lampiran 10 Pre-Processing Data

```
PRE-PROCESSING DATA

CASE FOLDING

def case_folding(text):
    # Mengubah teks menjadi huruf kecil
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text

df['case_folding'] = df['review'].apply(case_folding)
df.head(5)

Tampilkan output tersembunyi

Langkah berikutnya:  

CLEANING

[47] def clean_text(tweet):
    tweet = re.sub(r'https?:\V\S+', '', tweet)
    tweet = re.sub(r'[a-zA-Z]', '', tweet)

    emoji_pattern = re.compile("[
        u'\U0001F600-\U0001F64F' # emoticons
        u'\U0001F300-\U0001F5FF' # symbols & pictographs
        u'\U0001F680-\U0001F6FF' # transport & map symbols
        u'\U0001F1E0-\U0001F1FF' # flags (iOS)
        ]+", flags=re.UNICODE)
    tweet = emoji_pattern.sub(r'', tweet)

    tweet = re.sub(r'\d+', '', tweet)
    tweet = re.sub(r'\b\w\b', '', tweet)
    tweet = re.sub(r'\s+', ' ', tweet).strip()

    return tweet

df['cleansing'] = df['case_folding'].apply(clean_text)
df.head(5)
```

```
TOKENISASI

def tokenize(text):
    return word_tokenize(text)

df['tokenizing'] = df['cleansing'].apply(tokenize)
df.head(5)

Tampilkan output tersembunyi

Langkah berikutnya:  

STOPWORD

[49] import pandas as pd
from nltk.corpus import stopwords
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Assuming the dataframe 'df' and column 'token' already exist

# Define stopwords
stop_words = stopwords.words('indonesian')

additional_stopwords = [
    'yg','bi','d','g','nya','ya','sih','aku','kamu','dia','mereka',
    'kita','saya','kalian','ini','itu','loh','dong','kan','uh','aa',
    'uh','kak','ah','oh','hmm','eh','aduh','ad','deh','hehe','ehh',
    'gk','h','hh','huh','udah','mau','gw','lu','lo','loh','pake','ntah',
    'ma','km','kalo','kl','tuh','set','dll','mcm','mm','gan','bro','sis','de',
    'btw','kayak','spt','bilang','kata','tau','lihat','bikin','dan','atau',
    'setelah','juga','ts','btul','tg','tr','utk','kmrn','tdk','sy','sm','dm',
    'hrs','dg','dl','dt','sdh','tdk','tp','bsa','muas',
]

stop_words.extend(additional_stopwords)

def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

df['stopword_removal'] = df['tokenizing'].apply(lambda x: remove_stopwords(x))
df.head(5)
```

## STEMMING

```
[50] from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

# Membuat stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stem_text(tokens):
    # Mengembalikan kata dasar untuk setiap token
    return [stemmer.stem(word) for word in tokens]

# Terapkan fungsi stem_text ke kolom 'stopword_removal' dan gabungkan kembali menjadi string
df['stemming'] = df['stopword_removal'].apply(stem_text)
df['stemming'] = df['stemming'].apply(lambda x: ' '.join(x))

# Tampilkan 5 baris pertama dari DataFrame
df.head(5)
```

## Lampiran 11 Pembobotan Kata TF-IDF

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer

# Memuat data yang telah diproses sebelumnya
documents = df['stemming'].tolist()
N = len(documents)

# Inisialisasi TF-IDF Vectorizer
vectorizer = TfidfVectorizer()

# Melakukan fit dan transformasi pada dokumen
tfidf_matrix = vectorizer.fit_transform(documents)

# Mendapatkan nama fitur (terms)
terms = vectorizer.get_feature_names_out()

# Mendapatkan nilai IDF
idf_values = vectorizer.idf_

# Mengonversi matriks TF-IDF menjadi DataFrame
tfidf_df = pd.DataFrame(tfidf_matrix.toarray(), columns=terms)

# Menggabungkan DataFrame TF-IDF dengan label sentimen
final_df = pd.concat([tfidf_df, df['sentimen']], axis=1)

# Mencetak 500 baris pertama dari DataFrame akhir
print(final_df.head(500))

Tampilkan output tersembunyi

AMPILKAN KATA-KATA YANG SERING MUNCUL

# Ambil daftar kata kunci (terms)
terms = vectorizer.get_feature_names_out()

# Menghitung frekuensi setiap istilah dalam seluruh dataset
term_frequencies = tfidf_matrix.sum(axis=0).A1
terms_frequencies_df = pd.DataFrame({'term': terms, 'frequency': term_frequencies})

# Mengurutkan istilah berdasarkan frekuensi dalam urutan menurun
sorted_terms = terms_frequencies_df.sort_values(by='frequency', ascending=False)

# Menampilkan istilah yang paling sering muncul
print("Most frequent terms:")
print(sorted_terms.head(10))
```

## Lampiran 12 Split Data dan Evaluasi Confusion Matrix Metode KNN

```
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score

# Membagi data menjadi set pelatihan dan pengujian (80% pelatihan, 20% pengujian)
X_train, X_test, y_train, y_test = train_test_split(final_df.drop('sentimen', axis=1),
                                                  final_df['sentimen'],
                                                  test_size=0.3,
                                                  random_state=42)

# Menginisialisasi classifier KNN dengan parameter yang disesuaikan
knn_classifier = KNeighborsClassifier(n_neighbors=5) # Sesuaikan jumlah neighbors sesuai kebutuhan

# Melatih classifier
knn_classifier.fit(X_train, y_train)

# Prediksi pada set pengujian
y_pred = knn_classifier.predict(X_test)

# Akurasi
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Menghitung dan memvisualisasikan confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(10, 7))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=['POSITIF', 'NEGATIF'], yticklabels=['POSITIF', 'NEGATIF'])
plt.xlabel('Prediksi')
plt.ylabel('Actual')
plt.title('Confusion Matrix (KNN)')
plt.show()
```

## Lampiran 13 Hasil Akurasi Pada Nilai K

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score

# Mencari akurasi untuk berbagai nilai k
k_values = range(1, 21)
accuracies = []

for k in k_values:
    knn_classifier = KNeighborsClassifier(n_neighbors=k)
    knn_classifier.fit(X_train, y_train)
    y_pred = knn_classifier.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    accuracies.append((k, accuracy))

# Print hasil akurasi untuk setiap nilai k
print("Akurasi untuk berbagai nilai K:")
for k, accuracy in accuracies:
    print(f"Nilai K: {k}, Akurasi: {accuracy:.4f}")
```

## Lampiran 14 Plot Akurasi Nilai K

```
# Mencari akurasi untuk berbagai nilai k
k_values = range(1, 21)
accuracies = []

for k in k_values:
    knn_classifier = KNeighborsClassifier(n_neighbors=k)
    knn_classifier.fit(X_train, y_train)
    y_pred = knn_classifier.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    accuracies.append(accuracy)

# Plotting akurasi terhadap nilai k
plt.figure(figsize=(10, 6))
plt.plot(k_values, accuracies, marker='o')
plt.title('Akurasi terhadap Nilai K')
plt.xlabel('Nilai K')
plt.ylabel('Akurasi')
plt.xticks(k_values)
plt.grid(True)
plt.show()
```

## Lampiran 15 Visualisasi *Wordcloud* Positif dan Negatif

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Menginisialisasi TF-IDF Vectorizer
vectorizer = TfidfVectorizer()

# Menyesuaikan dan mengubah dokumen
tfidf_matrix = vectorizer.fit_transform(documents)

# Mendapatkan nama fitur (istilah)
terms = vectorizer.get_feature_names_out()

# Membuat DataFrame dari matriks TF-IDF
tfidf_df = pd.DataFrame(tfidf_matrix.toarray(), columns=terms)

# Menggabungkan DataFrame TF-IDF dengan label sentimen
final_df = pd.concat([tfidf_df, df['sentimen']], axis=1)

# Membuat DataFrame terpisah untuk setiap kategori sentimen
sentiments = df['sentimen'].unique() # Mendapatkan kategori sentimen yang unik

# Membuat fungsi untuk menghasilkan dan menampilkan WordCloud
def generate_wordcloud(text, title):
    wordcloud = WordCloud(width=800, height=400, background_color='white').generate_from_frequencies(text)
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.title(title)
    plt.show()

# Menghasilkan WordCloud untuk setiap kategori sentimen
for sentiment in sentiments:
    # Memilih baris berdasarkan sentimen
    subset_df = final_df[final_df['sentimen'] == sentiment]

    # Menghitung frekuensi kata (jumlah nilai TF-IDF per kolom)
    word_freq = subset_df.drop(columns=['sentimen']).sum(axis=0).to_dict()

    # Menghasilkan dan menampilkan WordCloud
    generate_wordcloud(word_freq, f"WordCloud - sentimen {sentiment}")
```

## Lampiran 16 Kartu Kendali Bimbingan

### KARTU KENDALI BIMBINGAN

Nama Mahasiswa : Ikhsan Nuttakwa Takbirata Ihram Nabawi  
NIM : 2011102441072  
Nama Dosen Pembimbing : Rudiman, S.Kom, M.Sc  
Judul Penelitian : Analisis Sentimen pada Ulasan Aplikasi Google Maps terhadap Pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Samarinda Menggunakan Metode K-Nearest Neighbor dengan Fitur Ekstraksi TF-IDF

NO	TANGGAL	URAIAN PEMBIMBINGAN	PARAF DOSEN
1	7 Feb 2024	Bimbingan pertama mendiskusikan persetujuan bimbingan dengan dosen Bapak Rudiman S.Kom, M.Sc. dengan topik penelitian Data Sentiment Analyst	
2	14 Feb 2024	Mencari permasalahan topik yang akan diteliti untuk dijadikan judul	
3	22 Feb 2024	Melakukan tahap pencarian data menggunakan python di Google Colab	
4	29 Feb 2024	Menentukan judul dan mulai perencanaan latar belakang dengan syarat ketentuan yang telah ditetapkan oleh dosen pembimbing, yaitu, menggunakan sumber jurnal paling sedikit 13 dan maksimal 15 jurnal untuk BAB 1 dan bersinta minimal sinta 4	
5	9 Maret 2024	Dalam pengerjaan latar belakang harus mengikuti aturan yang telah ditetapkan oleh dospem	
6	13 Maret 2024	Melakukan beberapa revisi terhadap mahasiswa yang dibimbing	
7	18 Maret 2024	Memberikan beberapa saran terhadap canvas pengajuan judul ke prodi	
8	27 Maret 2024	Memberikan beberapa revisi terhadap latar belakang rumusan masalah batasan masalah tujuan penelitian dan manfaat penelitian	

9	5 April 2024	Masuk BAB II menggunakan strategi yang telah diberikan oleh dosen pembimbing menggunakan pembobotan TF-IDF dalam penelitian yang dilakukan	
10	9 April 2024	Memberi masukan dan revisi dan bab II	
11	24 April 2024	Membahas target penelitian dan mencari sumber jurnal serta memberikan revisian pada BAB I dan BAB II	

Dosen Pembimbing



Rudjiman, S.Kom, M.Sc  
NIDN. 1105068202

Mengetahui,  
Ketua Program Studi



Arbansvah, S.Kom., M.TI  
NIDN. 1118019203

# SKRIPSI IKHSAN NUTTAKWA TAKBIRATA IHRAM NABAWI

by S1 Teknik Informatika Universitas Muhammadiyah Kalimantan Timur



---

Submission date: 25-Jul-2024 09:24AM (UTC+0800)

Submission ID: 2422037353

File name: SKRIPSI\_IKHSAN\_NUTTAKWA\_TAKBIRATA\_IHRAM\_NABAWI.docx (1.03M)

Word count: 5312

Character count: 33815

# SKRIPSI IKHSAN NUTTAKWA TAKBIRATA IHRAM NABAWI

*Ail*

## ORIGINALITY REPORT

**29%**  
SIMILARITY INDEX

**26%**  
INTERNET SOURCES

**14%**  
PUBLICATIONS

**12%**  
STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://repository.dinamika.ac.id">repository.dinamika.ac.id</a> Internet Source	2%
2	<a href="http://repository.uin-suska.ac.id">repository.uin-suska.ac.id</a> Internet Source	1%
3	<a href="http://ejournal.unesa.ac.id">ejournal.unesa.ac.id</a> Internet Source	1%
4	<a href="http://jurnal.polsri.ac.id">jurnal.polsri.ac.id</a> Internet Source	1%
5	<a href="http://docplayer.info">docplayer.info</a> Internet Source	1%
6	<a href="http://www.researchgate.net">www.researchgate.net</a> Internet Source	1%
7	<a href="http://123dok.com">123dok.com</a> Internet Source	1%
8	<a href="http://jurnal.kaputama.ac.id">jurnal.kaputama.ac.id</a> Internet Source	1%
9	<a href="http://doaj.org">doaj.org</a> Internet Source	1%

## RIWAYAT HIDUP



Ikhsan Nuttakwa Takbirata Ihram Nabawi, lahir di samarinda pada tanggal 09 November 2001. Merupakan anak kedua dari dua bersaudara yang lahir dari pasangan Bapak Masrul dan Ibu Erna M. memiliki saudari bernama Dina Mi'ratul Jannata Rayyana Nabawiyah. Pendidikan yang telah di tempuh oleh peneliti yaitu MI Sullamul Ulum Samarinda lulus pada tahun 2014. Kemudian dilanjutkan dengan menempuh Pendidikan di SMP Negeri 16 Samarinda lulus pada tahun 2017. Setelah itu dilanjutkan dengan menempuh Pendidikan di SMA Negeri 8 Samarinda, lulus pada tahun 2020. Pada tahun 2020 peneliti menempuh Pendidikan Program Studi S1 Teknik Informatika di Universitas Muhammadiyah Kalimantan Timur sebagai syarat kelulusan dan memperoleh gelar Sarjana Komputer (S.Kom).