# BAB II
# METODELOGI PENELITIAN

## 2.1. Objek Penelitian

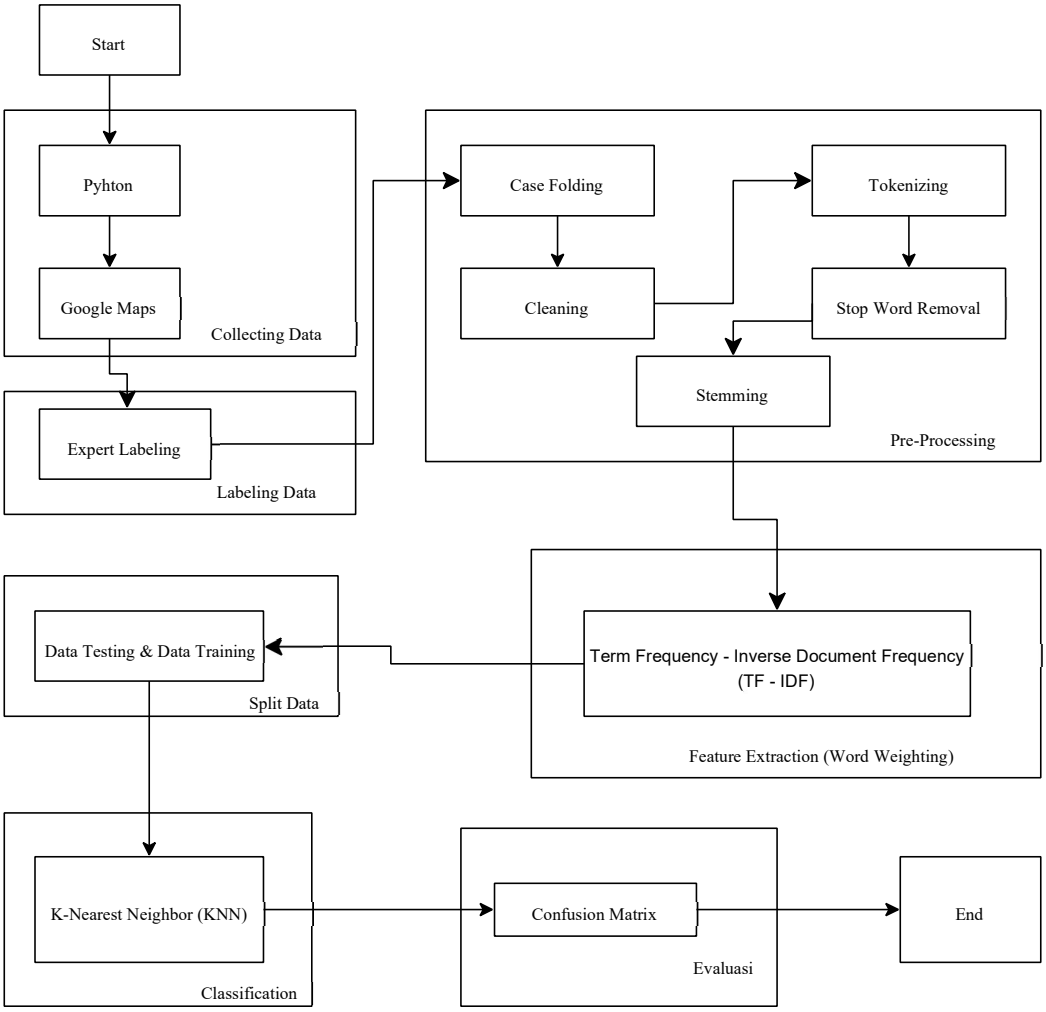
Objek penelitian dalam skripsi ini adalah ulasan pengguna Google Maps terhadap layanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Samarinda. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat yang menggunakan layanan BPJS Kesehatan tersebut dengan menggunakan algoritma *K-Nearest 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.

```
            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 beberapa sering kata tersebut muncul dalam dokumen tersebut. (Fadiyah Basar et al., 2022).

Penggunaan TF dapat menggunakan rumus pada Persamaan (2.1).

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

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

$$idf_t = log\frac{N}{df_t} \tag{2.2}$$

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).

$$W_{t,d} = tf_{d,t}\,X\,idf_t \tag{2.3}$$

Keterangan :

t = Kata kunci, term

d = Dokumen

W d,t = 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).

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

Keterangan :

pi = sample data / *data training*

qi = 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).

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

   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 (*Crawling*) | | | | | | |
| 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 | | | | | | |