

BAB III

METODE PENELITIAN

Bab ini menjelaskan pendekatan, desain, serta tahapan-tahapan dalam pelaksanaan penelitian. Penelitian ini dilakukan menggunakan pendekatan kuantitatif berbasis data digital, dengan memanfaatkan data media sosial untuk dianalisis melalui pendekatan *Natural Language Processing (NLP)*, dan *Graph Network Analysis (GNA)*.

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif komputasional yang bertujuan untuk memahami pola dan struktur reaksi publik terhadap Putusan Mahkamah Konstitusi Nomor 60/PUU-XXII/2024 melalui analisis data besar (*big data*) dan *text mining* yang diperoleh dari media sosial. Pendekatan kuantitatif dipilih untuk memungkinkan peneliti melakukan pengukuran sistematis terhadap sentimen dan hubungan antaraktor atau simpul komunikasi dalam jaringan diskursus publik. Sementara itu, metode eksploratif komputasional memberikan kerangka kerja yang relevan untuk menggali fenomena sosial secara luas dan mendalam dengan memanfaatkan teknik-teknik pengolahan data digital berbasis algoritma dan pemrograman. Data akan dikumpulkan secara **web crawling** dan **API streaming** dari media sosial dengan kata kunci yang berkaitan dengan putusan MK, seperti “MK 60/PUU”, “putusan MK pilkada”, dan “ambang batas pilkada”. Setelah data dikumpulkan, penelitian ini akan melakukan dua tahap analisis utama:

a) Analisis Sentimen

Sentimen publik dalam penelitian ini akan dibagi ke dalam tiga klasifikasi utama, yaitu positif, negatif, dan netral. Untuk mengidentifikasi kecenderungan emosi serta sikap masyarakat terhadap putusan Mahkamah Konstitusi tersebut, akan digunakan teknik *Natural Language Processing (NLP)*. Model analisis sentimen akan dikembangkan dan dilatih dengan memanfaatkan algoritma *Naïve Bayes Classifier*;

b) Analisis Jaringan (*Graph Network Analysis*)

Penelitian ini juga akan membangun jaringan interaksi sosial untuk

memetakan pola komunikasi dan identifikasi sumber-sumber yang menjadi kunci (*key opinion*) dipublik. Kemudian dianalisis dengan pendekatan *Social Network Analysis* (SNA) menggunakan metrik seperti *degree centrality*, *betweenness centrality*.

Dengan pendekatan ini, diharapkan penelitian dapat memberikan gambaran kuantitatif dan struktural mengenai bagaimana reaksi publik terbentuk, tersebar, dan dimobilisasi melalui media sosial pasca putusan MK tersebut. Selain itu, penelitian ini juga berkontribusi pada pengembangan metodologi analisis politik berbasis data digital di Indonesia, khususnya dalam memahami dinamika opini publik terhadap isu-isu konstitusional dan demokrasi.

3.2 Desain Penelitian

Desain penelitian dalam implementasi penulisan tugas akhir ini mencakup langkah-langkah yang penting dalam memastikan bahwa penelitian dilakukan dengan metodologi yang tepat dan relevan dengan tema yang dipilih. Penelitian ini menggunakan desain studi observasional non-eksperimental, di mana peneliti mengumpulkan data dari media sosial dan berita *online* tanpa melakukan intervensi langsung terhadap subjek. Berikut tahapan-tahapan yang ditunjukkan untuk melakukan penelitian pada gambar 3.1.



Gambar 3. 1 Tahapan Desain Penelitian

3.3 Subjek Penelitian

Subjek penelitian adalah konten media sosial publik yang membahas Putusan Mahkamah Konstitusi Nomor. 60/PUU-XXII/2024. Fokus diutamakan pada *platform* Sosial media dan Berita *Online* pada bagian kolom komentar yang disediakan untuk pembaca dan penonton mengingat keterbukaan data dan format teks pendek yang cocok untuk teknik *Natural Language Processing* (NLP). Pada penelitian ini penulis memilih 3 kata kunci atau *keyword* yang digunakan untuk dasar pencarian sumber data untuk diolah adalah “MK 60/PUU”, “putusan MK pilkada” dan “ambang batas pilkada”.

3.4 Instrumen Penelitian

Instrumen utama yang digunakan dalam penelitian ini meliputi:

- 1) **Perangkat Lunak dan Pustaka Python**
 - a) *API Straming*, untuk mengakses data secara langsung dan *real-time* dari suatu layanan atau *platform*—biasanya berupa data yang terus mengalir seperti postingan media sosial, data sensor, harga pasar, atau pesan dari sistem.
 - b) *KNIME*: untuk analisis sentimen menggunakan algoritma pembelajaran mesin;
- 2) **Model Klasifikasi Sentimen**: Algoritma Naïve Bayes *Classifier* yang dilatih menggunakan data set teks berbahasa Indonesia.
- 3) **Tahapan Pra-pemrosesan Teks**:
 - a) **Data Cleaning**, *Data cleaning* adalah langkah data *preprocessing* untuk meningkatkan kualitas data dengan menghapus atau mengoreksi informasi yang salah, tidak lengkap, atau tidak relevan. Tahapan ini mencakup pendeteksian dan perbaikan masalah data, menghapus kolom yang tidak diinginkan, dan menyelesaikan ketidaksesuaian untuk memastikan data yang lebih bersih dan akurat (Gharatkar et al., 2017);
 - b) **Case Folding** : Proses konversi seluruh huruf kapital dalam teks menjadi huruf kecil agar seragam dalam pemrosesan data;
 - c) **Stopwords Removal** : Tahapan menghapus kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis sentimen, seperti “aku”, “dan”, “atau”, “adalah”, dan sejenisnya;
 - d) **Text Normalization** : Proses penyesuaian ejaan yang salah atau kata tidak baku menjadi bentuk standar, termasuk menghilangkan kata-kata tidak relevan seperti “yg”, “gak”, “sdh”, “tp”, dan sebagainya;
 - e) **Stemming** : Teknik untuk mengembalikan kata turunan ke bentuk dasarnya. Misalnya, kata “menghormati” akan diubah menjadi “hormat”

- f) **TF-IDF vectorizer**, Metode untuk mengubah teks menjadi bentuk numerik berbasis bobot kata, sehingga dapat digunakan dalam proses analisis atau klasifikasi.

3.5 Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui API Twitter dengan memanfaatkan sejumlah kata kunci yang berkaitan dengan Putusan MK No. 60/PUU-XXII/2024, seperti:

- a) MK 60/PUU;
- b) Putusan MK pilkada;
- c) Ambang batas pilkada.

Pengumpulan data dilakukan mulai dari tanggal pengumuman putusan hingga satu minggu setelahnya, untuk menangkap dinamika opini secara menyeluruh. Kriteria data yang dikumpulkan adalah sebagai berikut:

- a) Bahasa Indonesia;
- b) Berasal dari akun publik (tidak terkunci/*protected*);
- c) Relevan dengan topik berdasarkan kata kunci yang ditentukan.

3.6 Teknik Analisis Data

3.6.1 Analisis Sentimen

Setelah proses melewati tahapan Pra Pemrosesan Teks dilakukan terhadap unggahan-unggahan media sosial yang memuat kata kunci dan tagar relevan dengan *keyword* yang telah ditetapkan “MK 60/PUU”, “putusan MK pilkada” dan “ambang batas pilkada”, langkah selanjutnya dalam penelitian ini adalah analisis sentimen menggunakan algoritma Naïve Bayes *Classifier*. Menurut Sebastiani (2002) dalam *Machine Learning in Automated Text Categorization*, Naïve Bayes *Classifier* merupakan algoritma klasifikasi probabilistik berbasis Teorema Bayes yang bekerja secara efektif dalam domain teks karena kesederhanaannya, kecepatan komputasi, dan performa yang cukup tinggi untuk tugas-tugas klasifikasi berbasis dokumen pendek, seperti *tweet* atau unggahan media sosial. Hasil dikodekan dan divisualisasikan untuk melihat distribusi opini secara kuantitatif.

3.6.2 Graph Network Analysis

Centrality digunakan sebagai indikator untuk mengidentifikasi aktor yang memiliki peran paling berpengaruh dalam suatu jaringan sosial, dengan menunjukkan tingkat keterpusatan posisi seseorang dalam jaringan (Mincer & Niewiadomska-Szynkiewicz, 2012). Terdapat empat jenis utama pengukuran *centrality*, yakni: *degree centrality*, *betweenness centrality*, *closeness centrality*, dan *eigenvector centrality*.

Tabel 3. 1 Pengertian *Centrality*

<i>Centrality</i>	Penjelasan
<i>Degree Centrality</i>	<i>Degree centrality</i> dihitung berdasarkan jumlah hubungan langsung (<i>edges</i>) yang terhubung ke suatu simpul (<i>node</i>) dalam jaringan (Alhajj & Rokne, 2014).
<i>Betweenness Centrality</i>	<i>Betweenness centrality</i> digunakan untuk mengidentifikasi simpul yang berfungsi sebagai jalur penghubung atau perantara dalam distribusi informasi antar simpul lain (Mbaru & Barnes, 2017).
<i>Closeness Centrality</i>	<i>Closeness centrality</i> mengukur seberapa dekat suatu simpul terhadap semua simpul lainnya di dalam jaringan, yang ditentukan dari rata-rata jarak terpendek ke seluruh <i>node</i> (Alhajj & Rokne, 2014).
<i>Eigenvector Centrality</i>	<i>Eigenvector centrality</i> menilai pentingnya sebuah simpul berdasarkan kualitas koneksi yang dimilikinya, termasuk keterkaitannya dengan simpul-simpul berpengaruh lainnya dalam jaringan (Alhajj & Rokne, 2014).

3.7 Prosedur Penelitian

Prosedur penelitian ini terdiri dari serangkaian tahapan sistematis yang rangkaiannya sama dengan diagram yang pada desain penelitian, berikut untuk penjelasan lengkapnya sebagai berikut:

a) **Pengumpulan Data Menggunakan Web Crawling**

Menentukan kata kunci seperti “*MK 60/PUU*”, “*putusan MK pilkada*”, dan frasa relevan lainnya. Rentang waktu pengambilan data dimulai dari hari pengumuman putusan yakni 20 Agustus 2024 hingga 1 minggu setelahnya yakni 27 Agustus 2024 untuk menangkap dinamika reaksi publik yang terdokumentasi pada kanal berita *online* dan media sosial;

b) **Data Selection**

Menurut **Han, Kamber, dan Pei (2012)** dalam bukunya *Data Mining: Concepts and Techniques*, data selection adalah: proses pengambilan data yang relevan dari basis data untuk digunakan dalam tugas data mining, dengan memperhatikan kesesuaian terhadap tujuan analisis dan kualitas data;

c) **Data Pre-Processing**

Data preprocessing merupakan langkah kedua setelah data selection yang sangat krusial dan penting dalam pengolahan data. Pada tahap ini, data dipersiapkan untuk analisis dengan tujuan menghilangkan data yang kurang relevan dan tidak konsisten (Gharatkar et al., 2017);

d) **Analisis Sentimen**

Analisis sentimen adalah suatu teknik yang digunakan untuk mengenali bagaimana opini atau perasaan tertentu tercermin dalam teks, serta bagaimana opini tersebut dapat dikelompokkan ke dalam kategori sentimen positif atau negatif (Nasukawa & Yi, 2003 dalam Faturohman 2023). Kemudian menurut pendapat yang dikemukakan oleh Cvijikj & Michahelles, di mana analisis sentimen digunakan untuk memahami komentar yang diciptakan oleh pengguna (internet) dan menjelaskan bagaimana sebuah produk maupun brand diterima oleh mereka (Cvijikj & Michahelles, 2011). Sedangkan menurut pendapat lain, analisis sentimen adalah proses yang digunakan untuk menentukan opini, emosi dan sikap yang dicerminkan melalui teks, dan biasanya diklasifikasikan menjadi opini negatif dan positif (Coletta et al., 2014). Berdasarkan ketiga pandangan tersebut, dapat disimpulkan bahwa analisis sentimen merupakan suatu proses yang bertujuan untuk mengidentifikasi dan

memahami opini atau perasaan yang dinyatakan oleh individu melalui teks, dengan kemampuan untuk mengklasifikasikan sentimen tersebut sebagai positif, negatif, atau netral;

e) **Graph Network Analysis**

Pengukuran *centrality* digunakan untuk menentukan aktor yang berperan paling penting dalam suatu jaringan sosial, hal ini menunjukkan derajat pusat seseorang (Mincer dan Niewiadomska-Szynkiewicz, 2012);

f) **Visualisasi**

Hasil dari ketiga pendekatan dianalisis dan dikaitkan dengan konteks sosial-politik dari Putusan MK. Laporan disusun dalam bentuk narasi ilmiah yang mencakup temuan, visualisasi, dan interpretasi mendalam.

3.8 Evaluasi dan Validasi Hasil

Untuk memastikan keandalan dan validitas hasil analisis, evaluasi dilakukan terhadap pada tiga aspek utama sebagai berikut:

3.8.1 Evaluasi Model Analisis Sentimen

Hasil olahan Model Naïve Bayes *Classifier* dapat dievaluasi untuk mengerahui tingkat akurasi dan efisiensinya, beberapa metrics yang harus diketahui dalam evaluasi model yaitu akurasi, F1-score, presisi dan *recall* (Tharwat, 2020). Akurasi merupakan tingkat ketepatan prediksi model terhadap label asli dari data yang diolah atau rasio kebenaran hasil prediksi terhadap label asli data tersebut.

- a) **Evaluation metrics**, merupakan tahap penelitian untuk menunjukkan kinerja atau kinerja akurasi suatu model. Sehingga model yang sudah dihasilkan melalui data set dan proses-proses sebelumnya akan dievaluasi menggunakan *metrics* atau yang biasa dikenal dengan istilah *confusion matrix*. Dari sini dapat dilihat kinerja model yang sudah dihasilkan apakah baik atau sebaiknya;
- b) **Akurasi** adalah tingkat ketepatan prediksi model terhadap informasi yang dimasukkan. Dengan akurasi dapat dihitung sebagai berikut :

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

- c) **F1-score**, yang memiliki nilai maksimum 1 dan nilai minimum 0, merupakan rata-rata harmonik dari presisi dan **recall**;

$$\frac{1}{F1\ Score} = \frac{1}{2} \left(\frac{1}{Presisi} + \frac{1}{Recall} \right)$$

- d) **Presisi** adalah nilai rasio klasifikasi yang akurat terhadap sasaran dengan seberapa banyak informasi yang dikelompokkan ke dalam kelas yang sama;

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

- e) **Recall** adalah rasio informasi yang tepat sasaran dibandingkan dengan semua informasi dengan label yang sama;

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

- f) **Splitting** (Pembagian *Data Training* dan *Data Testing*) Proses *splitting* adalah tahapan pembagian dataset ke dalam dua kelompok, yaitu data latih (*training data*) dan data uji (*testing data*). Data *training* digunakan untuk membangun dan melatih model klasifikasi, sedangkan data testing berfungsi untuk menguji performa model yang telah dilatih. Model klasifikasi dikembangkan berdasarkan kumpulan data latih, kemudian dievaluasi menggunakan data uji. Skema pembagian data umumnya mengikuti rasio tertentu, seperti 60:40, 70:30, 80:20, atau 90:10, yang masing-masing menunjukkan proporsi data untuk pelatihan dan pengujian. Perbandingan ini dilakukan untuk menentukan konfigurasi terbaik dalam menghasilkan tingkat akurasi tertinggi. Kualitas hasil klasifikasi sangat bergantung pada cakupan data training—semakin representatif data latih terhadap karakteristik data uji, maka semakin besar kemungkinan model menghasilkan akurasi yang optimal.

3.8.2 Validasi *Graph Network Analysis*

Validasi analisis jaringan dilakukan melalui tiga pendekatan:

- a) **Konsistensi Visualisasi**: Struktur jaringan yang dihasilkan harus mencerminkan dinamika diskusi yang logis dan sesuai dengan data interaksi;

- b) **Metrik Jaringan:** Metrik seperti *density* (kepadatan jaringan) dan *modularity* (tingkat pemisahan komunitas) digunakan untuk menilai kekuatan dan keterhubungan antar simpul;
- c) **Validasi Kontekstual:** akun sentral yang terdeteksi (berdasarkan metrik seperti *degree centrality* atau *betweenness centrality*) divalidasi untuk memastikan relevansi dalam konteks sosial Putusan MK Nomor. 60/PUU-XII/2024.

3.9 Timeline Penelitian

Berikut adalah *Timeline* Penelitian yang dijadikan acuan oleh penulis untuk menuntaskan proses penelitian kali ini,

Tabel 3. 2 *Timeline* Penelitian

Bulan	Kegiatan
Bulan 1	Penyusunan Proposal Penelitian - Penentuan topik dan rumusan masalah - Studi literatur dan penyusunan Alur Penelitian - Penyusunan metodologi penelitian
Bulan 2	Seminar Proposal dan Revisi - Presentasi proposal ke dosen pembimbing dan penguji - Revisi proposal berdasarkan masukan
Bulan 3	Pengumpulan dan Pra-Pemrosesan Data - crawling data media sosial dan berita online sesuai kata kunci - Data <i>cleaning</i> , normalisasi, dan seleksi data
Bulan 4	Analisis Data (Tahap I dan II) - Klasifikasi sentimen menggunakan algoritma Naïve Bayes - Pembangunan struktur <i>graph network</i> dan identifikasi relasi
Bulan 5	Interpretasi dan Pembahasan Hasil - Analisis kuantitatif terhadap persebaran opini publik - Analisis hubungan antar aktor berdasarkan struktur jaringan
Bulan 6	Finalisasi Skripsi dan Ujian - Revisi hasil pembimbingan - Uji plagiarisme - Ujian skripsi dan revisi pasca-sidang