

BAB III

METODOLOGI

3.1 Metodologi Penelitian

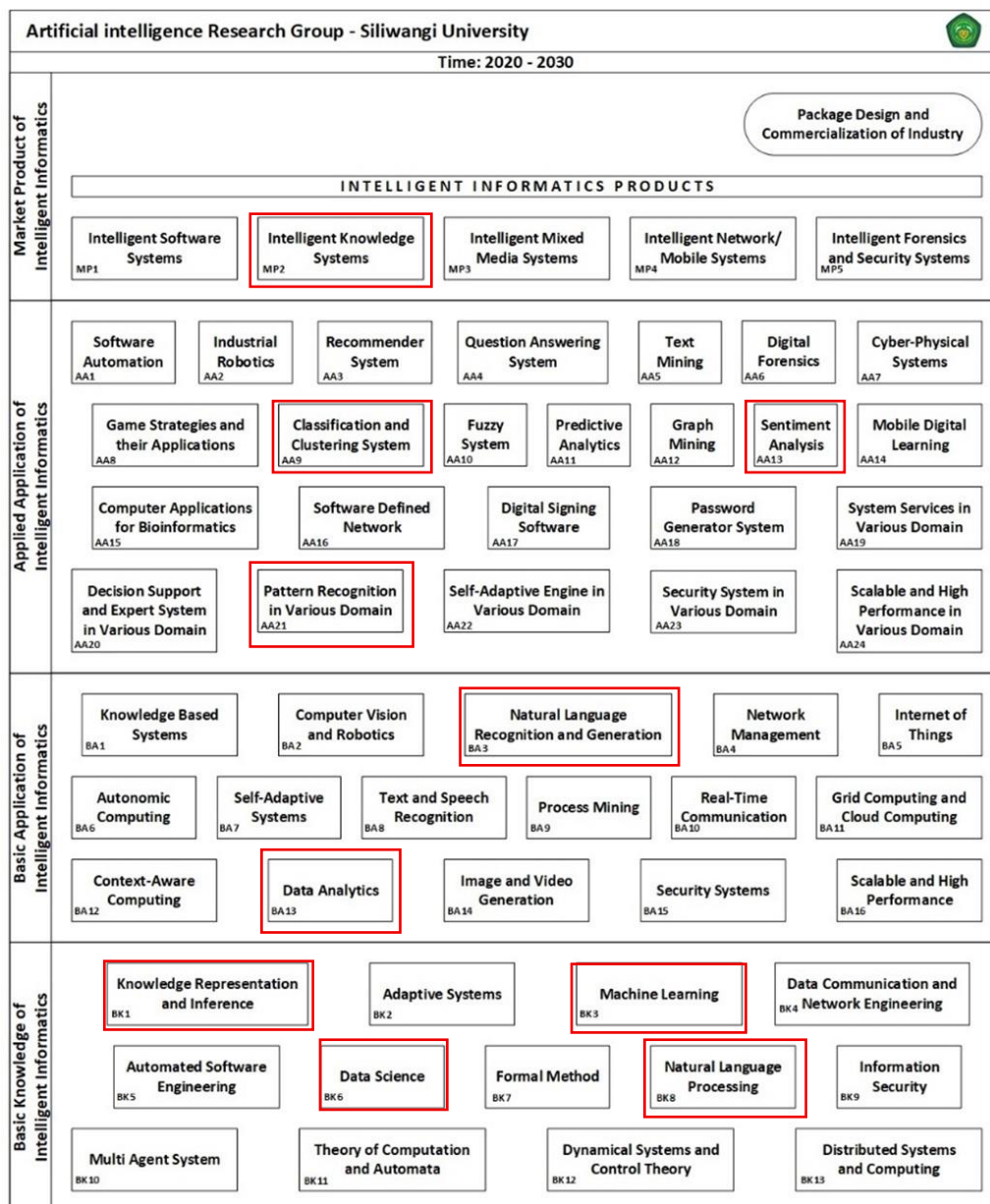
Penelitian ini menerapkan metode kuantitatif dengan pendekatan eksperimental. Pemilihan pendekatan kuantitatif didasarkan pada penggunaan data yang diolah melalui sistem komputasi dan dianalisis menggunakan perhitungan matematis, sehingga hasil analisis bersifat numerik atau statistik. Pendekatan eksperimen digunakan karena penelitian ini mengukur nilai dari masing-masing variabel, sehingga perlu dilakukan evaluasi terhadap pengaruh variabel independen terhadap variabel dependen (Creswell & Creswell, 2018).

Penelitian ini menggunakan objek berupa teks berita saham sektor perbankan berbahasa Indonesia yang dikumpulkan dari berbagai portal berita daring. Proses penelitian dilakukan melalui empat tahapan utama. Pertama, tahap pengumpulan data dilakukan dengan teknik *web scraping* untuk memperoleh korpus berita yang memuat informasi terkait pergerakan saham dan isu-isu di sektor perbankan. Kedua, tahap *preprocessing* dilakukan untuk membersihkan dan menyiapkan data. Tahap ketiga adalah pemodelan menggunakan *Aspect-Specific Graph Convolutional Network (ASGCN)* untuk membangun representasi graf dari hubungan antara aspek dan kata dalam teks, sehingga memungkinkan model memahami konteks sentimen terhadap aspek tertentu secara lebih efektif. Selain itu, untuk menilai keunggulan pendekatan berbasis aspek digunakan pula model pembandingan berupa baseline *Graph Convolutional Network (GCN)* dan *BiLSTM*.

Tahap terakhir adalah evaluasi model, di mana kinerja model dianalisis menggunakan metrik evaluasi seperti akurasi, *presisi*, *recall*, dan *F1-score* dalam tugas klasifikasi sentimen berbasis aspek terhadap berita saham sektor perbankan

3.2 Peta jalan (*Road Map*) Penelitian

Penelitian ini merupakan bagian dari *roadmap* penelitian KK ISI (Informatika dan Sistem Inteligen) Universitas Siliwangi. Kajian sebelumnya digunakan sebagai dasar dan sumber pengetahuan untuk menyempurnakan capaian akhir dari *roadmap* penelitian secara keseluruhan. Gambar 3.1 menjelaskan kajian penelitian yang akan dilakukan dalam penelitian ini, berfokus pada kajian yang ditandai di dalam garis persegi panjang berwarna merah, yaitu *Natural Language Processing*, *Machine Learning*, *Data Science*, *Knowledge Representation and Inference*, *Natural Language Recognition and Generation*, *Data Analytics*, *Sentiment Analysis*, *Classification and Clustering System*, *Pattern Recognition in Various Domain* dengan tujuan menghasilkan prototipe *Intelligent Knowledge System* tentang klasifikasi sentimen berbasis aspek. Penelitian ini berfokus untuk mengevaluasi efektivitas model *Aspect-Specific Graph Convolutional Network (ASGCN)* dalam melakukan analisis sentimen berbasis aspek terhadap berita saham pada sektor perbankan. Model ini diimplementasikan untuk mengidentifikasi dan mengklasifikasikan sentimen yang terkait dengan aspek-aspek spesifik dalam konten berita, sehingga diharapkan mampu memberikan pemahaman yang lebih mendalam mengenai persepsi pasar terhadap isu-isu tertentu dalam domain keuangan.

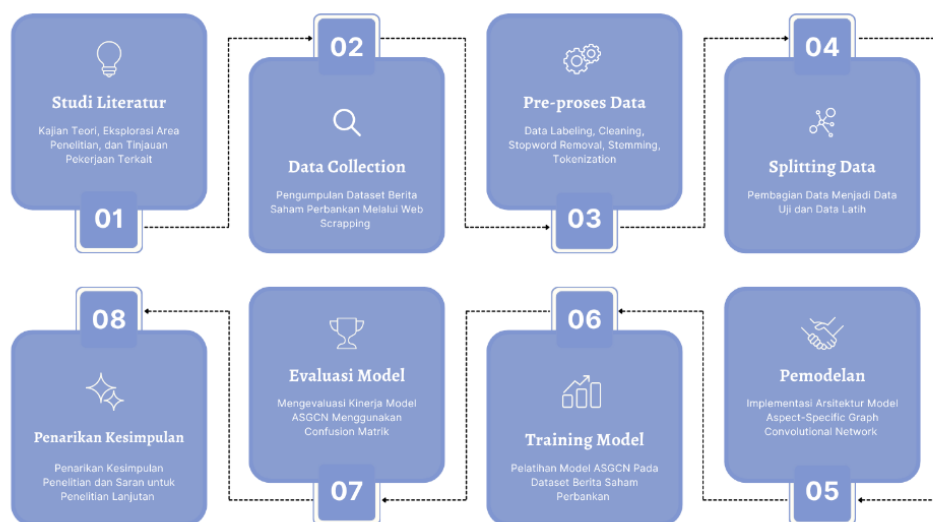


Gambar 3.1 Roadmap Penelitian (AIS Universitas Siliwangi, 2020)

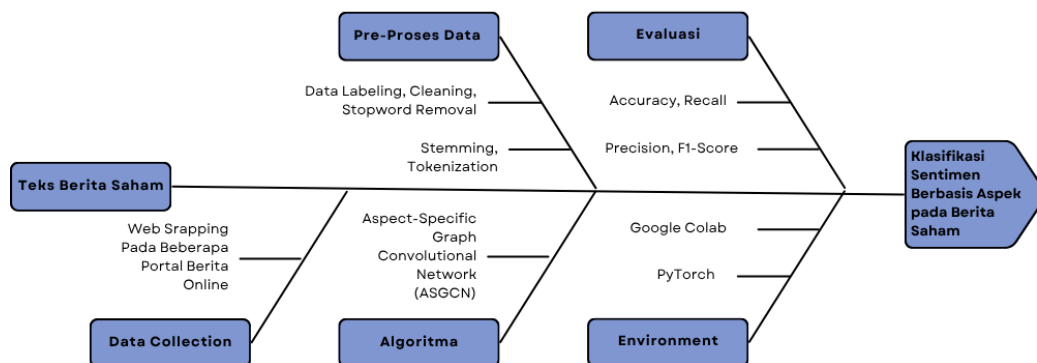
3.3 Tahapan Penelitian

Penelitian ini bertujuan menghadirkan kontribusi baru dalam bidang analisis sentimen terhadap berita saham dengan menerapkan klasifikasi sentimen berbasis aspek pada berita saham sektor perbankan. Pendekatan tersebut diimplementasikan melalui penggunaan model *Aspect-Specific Graph Convolutional Network*

(ASGCN) yang dirancang untuk menangkap hubungan antara aspek dan sentimen dalam teks secara lebih kontekstual dan terstruktur. Diharapkan, penerapan model *Aspect-Specific Graph Convolutional Network (ASGCN)* dalam penelitian ini tidak hanya meningkatkan efektivitas klasifikasi sentimen berbasis aspek, tetapi juga memberikan kontribusi terhadap pengembangan metode analisis sentimen dalam konteks berita saham. Tahapan penelitian meliputi studi literatur, pengumpulan data, *preprocessing* data, pemisahan data, pelatihan model, evaluasi model serta penarikan kesimpulan. Rangkaian tahapan penelitian secara rinci disajikan pada Gambar 3.2. Sementara itu, tujuan dari penelitian ini divisualisasikan dalam bentuk diagram fishbone, yaitu diagram yang menggambarkan berbagai faktor penyebab dari suatu fenomena spesifik (Project Management Institute., 2015), sebagaimana ditunjukkan pada Gambar 3.3.



Gambar 3.2 Tahapan Penelitian

Gambar 3.3 *Fishbond Diagram*

Bagian ekor diagram *fishbone* pada Gambar 3.3 mewakili objek penelitian, yaitu teks berita saham, sedangkan bagian kepala menunjukkan tujuan penelitian, yaitu klasifikasi sentimen berbasis aspek pada berita saham dengan menggunakan model *Aspect-Specific Graph Convolutional Network*. Sementara bagian tulang tengahnya menjelaskan elemen–elemen yang mendukung tercapainya tujuan dari penelitian, dengan keterangan sebagai berikut:

- a. *Dataset*, data dikumpulkan melalui proses *web scraping* dari beberapa portal berita online yang memuat informasi terkait saham. Proses ini bertujuan untuk memperoleh korpus teks yang akan digunakan sebagai dasar dalam pelatihan model klasifikasi sentimen berbasis aspek.
- b. *Pre-proses Data*, Data mentah yang telah dikumpulkan kemudian melalui serangkaian proses praproses untuk memastikan kualitas dan konsistensi data. Proses ini meliputi *Data labeling*, *cleaning*, dan *stopword removal* untuk membuang kata-kata yang tidak relevan, serta *Stemming* dan *tokenization* untuk mengubah kata ke bentuk dasar dan memecah teks menjadi token yang dapat diolah lebih lanjut oleh model.

- c. Algoritma, Penelitian ini menggunakan model *Aspect-Specific Graph Convolutional Network (ASGCN)* yaitu metode berbasis *Graph Neural Network* yang secara khusus dirancang untuk menangani klasifikasi sentimen berdasarkan aspek yang disebutkan dalam teks. Model ini mampu memahami hubungan antar kata dalam struktur kalimat serta memfokuskan perhatian pada aspek-aspek tertentu.
- d. *Environment*, Implementasi dan pelatihan model dilakukan pada platform *Google Colab* dengan menggunakan framework *PyTorch*. Pemilihan environment ini memungkinkan pemrosesan yang efisien dengan dukungan *GPU*, serta fleksibilitas dalam pengembangan dan eksperimen.
- e. Evaluasi, Kinerja model dikaji menggunakan berbagai metrik evaluasi seperti *Accuracy*, *Precision*, *recall* dan *F1-score* untuk memperoleh gambaran menyeluruh mengenai efektivitas klasifikasi.

3.3.1 Studi Literatur

Tahap ini bertujuan untuk mengeksplorasi bidang penelitian serta mengidentifikasi studi terdahulu yang relevan. Kajian teori dilakukan untuk memahami konsep dasar serta perkembangan terbaru dalam topik yang diteliti. Berdasarkan literatur yang telah dikaji, dilakukan identifikasi kesenjangan penelitian (*research gap*) sebagai dasar perumusan *Research Problems (RP)*. Selanjutnya, ditetapkan *Research Questions (RQ)* dan *Research Objectives (RO)* untuk merumuskan arah penelitian serta menentukan kontribusi akademik yang dapat diberikan.

3.3.2 Data Collection

Setelah ditetapkan arah penelitian, tahap selanjutnya adalah pengumpulan data. Data dikumpulkan melalui proses *web scraping* terhadap portal berita daring yang memuat informasi terkait saham, khususnya sektor perbankan. Hasil scraping ini menjadi dataset utama yang akan digunakan dalam proses pelatihan dan pengujian model klasifikasi sentimen berbasis aspek.

3.3.3 Pre-Proses Data

Data yang telah dikumpulkan kemudian diproses agar siap digunakan dalam pemodelan. Tahap pre-proses pada penelitian ini mencakup beberapa tahapan berikut:

- a. *Data Labeling*, Pelabelan data dilakukan dengan menggunakan *ABSA Dataset Prepare Tools* yang disediakan oleh perpustakaan *PyABSA* (Alifi dkk., 2023b). Sebelum anotasi, data dibagi menjadi potongan-potongan yang lebih kecil, karena alat anotasi memiliki keterbatasan dan tidak dapat membaca jumlah data yang besar. Setelah dilabeli, struktur data berubah menjadi format *Aspect Polarity Classification (APC)*. *APC* terdiri dari tiga kolom: teks, aspek, dan polaritas aspek. Aspek dalam teks berita akan ditandai dengan "\$T\$".
- b. *Cleaning*, bertujuan untuk menghilangkan elemen-elemen yang tidak diperlukan dalam teks. Langkah-langkah yang dilakukan dalam tahap ini meliputi menghapus karakter khusus seperti tanda baca, angka, dan simbol yang tidak relevan. Selain itu, proses ini juga mencakup penghapusan karakter berulang atau *whitespace* berlebih yang muncul dalam teks, serta

mengonversi teks ke huruf kecil (*lowercasing*) untuk menjaga konsistensi data.

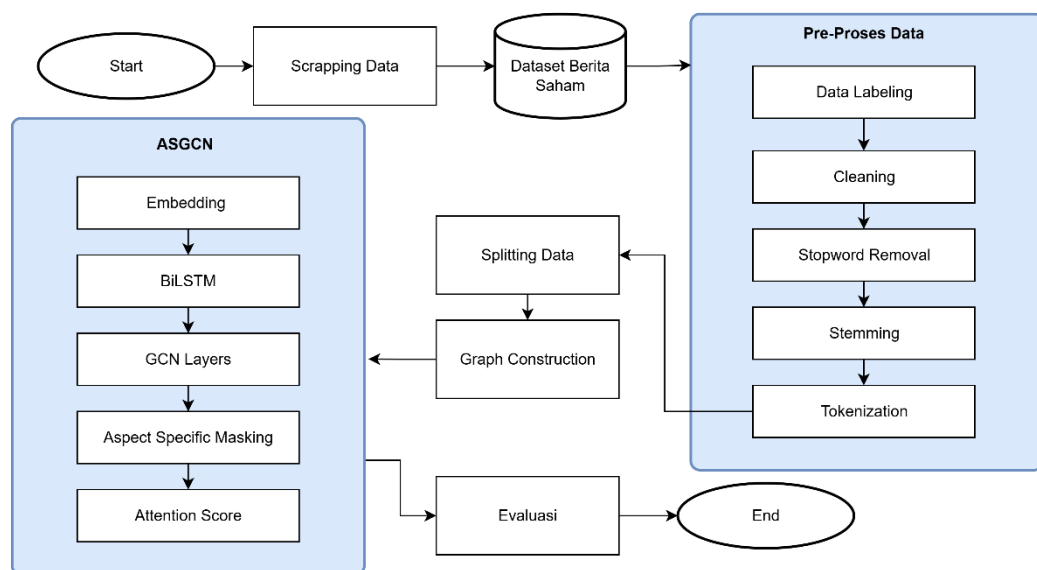
- c. *Stopword Removal*, menghilangkan kata-kata umum yang tidak memberikan makna signifikan dalam teks.
- d. *Stemming*, mengubah kata ke bentuk dasarnya berdasarkan makna dalam kamus.
- e. *Tokenization*, memecah teks menjadi bagian-bagian yang lebih kecil (token) yang dapat diproses oleh model.

3.3.4 *Splitting Data*

Data dibagi dengan rasio 80% untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Pemilihan rasio ini didasarkan pada standar umum (*rule of thumb*) dalam eksperimen pembelajaran mesin yang sering dikaitkan dengan Prinsip Pareto. Rasio ini dipilih karena dinilai memberikan keseimbangan yang optimal dimana 80% data memberikan variasi yang cukup bagi model untuk mempelajari pola dan fitur yang kompleks, sedangkan 20% sisanya dianggap representatif secara statistik untuk melakukan validasi yang objektif guna memastikan bahwa model dapat menggeneralisasi dengan baik terhadap data yang belum pernah dilihat selama pelatihan (Febrianto dkk., 2024). Data latih digunakan untuk melatih model dalam mengenali pola, sementara data uji digunakan untuk mengevaluasi performa model terhadap data baru yang belum pernah dilihat sebelumnya

3.3.5 Pemodelan

Tahap ini mencakup implementasi arsitektur model *Aspect-Specific Graph Convolutional Network (ASGCN)*. Model ini dirancang untuk memahami hubungan semantik antar kata dalam kalimat dan mengidentifikasi sentimen terhadap aspek-aspek tertentu dalam berita saham. Proses ini menjadi inti dari pendekatan klasifikasi sentimen berbasis aspek.



Gambar 3.4 Arsitektur Model

Berdasarkan Gambar 3.4 arsitektur model dalam penelitian ini diawali dengan proses pengumpulan data melalui teknik *web scraping* dari beberapa portal berita daring yang memuat informasi terkait berita saham. Data yang diperoleh kemudian melalui tahap pra-pemrosesan (*preprocessing*), yang terdiri atas beberapa langkah, yaitu *data labeling*, *cleaning*, *stopword removal*, *Stemming*, dan *tokenization*. Selanjutnya, dataset dibagi menjadi dua subset, yaitu 80% untuk data latih dan 20% untuk data uji. Setelah itu, dilakukan *graph construction*, yaitu pembangunan representasi graf dari teks yang merepresentasikan relasi antar kata dalam struktur kalimat. Representasi graf ini menjadi komponen esensial dalam model *Aspect-*

Specific Graph Convolutional Network (ASGCN). Proses pemodelan dilakukan menggunakan arsitektur *ASGCN* yang terdiri atas beberapa komponen utama, antara lain *word embedding* untuk merepresentasikan kata dalam bentuk vektor berdimensi tetap, *Bi-directional Long Short-Term Memory (BiLSTM)* untuk menangkap dependensi sekuensial dari dua arah, *Graph Convolutional Network (GCN)* untuk memproses informasi berbasis graf, *aspect-specific masking* untuk memfokuskan perhatian pada kata-kata yang relevan terhadap aspek yang dianalisis, serta *attention mechanism* untuk memberikan bobot penting pada token yang paling berkontribusi terhadap prediksi sentimen. Tahap terakhir adalah evaluasi, yang bertujuan untuk mengukur performa dan efektivitas model *ASGCN* dalam melakukan klasifikasi sentimen berbasis aspek pada berita saham.

3.3.6 Training Model

Setelah arsitektur model diterapkan, dilakukan pelatihan model menggunakan data latih yang telah dipersiapkan. Pelatihan ini bertujuan untuk mengoptimalkan bobot pada jaringan saraf agar model mampu mengenali pola-pola sentimen yang berkaitan dengan aspek-aspek tertentu dalam teks berita saham perbankan.

3.3.7 Evaluasi Model

Model yang telah dilatih kemudian dievaluasi menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Penilaian juga dilakukan dengan menggunakan *confusion matrix* untuk melihat distribusi hasil prediksi model

terhadap data uji, serta untuk mengidentifikasi seberapa baik model dalam mengklasifikasikan sentimen secara tepat terhadap tiap aspek.

3.3.8 Penarikan Kesimpulan

Tahap akhir dalam penelitian ini adalah penarikan kesimpulan berdasarkan hasil evaluasi model. Kesimpulan mencakup efektivitas model *ASGCN* dalam melakukan klasifikasi sentimen berbasis aspek pada berita saham. Selain itu, diberikan pula saran dan rekomendasi untuk penelitian selanjutnya yang dapat mengembangkan metode ini lebih lanjut atau menerapkannya pada domain lain.