

BAB III

METODOLOGI

3.1 Metodologi Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen. Pendekatan kuantitatif dipilih karena penelitian ini bersumber dari data yang telah diuji menggunakan sistem komputasi serta diterapkan perhitungan matematis, sehingga analisis akhirnya bersifat statistik atau numerik. Pendekatan eksperimen diterapkan karena penelitian ini menghasilkan nilai untuk setiap variabel, sehingga diperlukan pengukuran terhadap pengaruh antara variabel independen dan dependen (Creswell & Creswell, 2018).

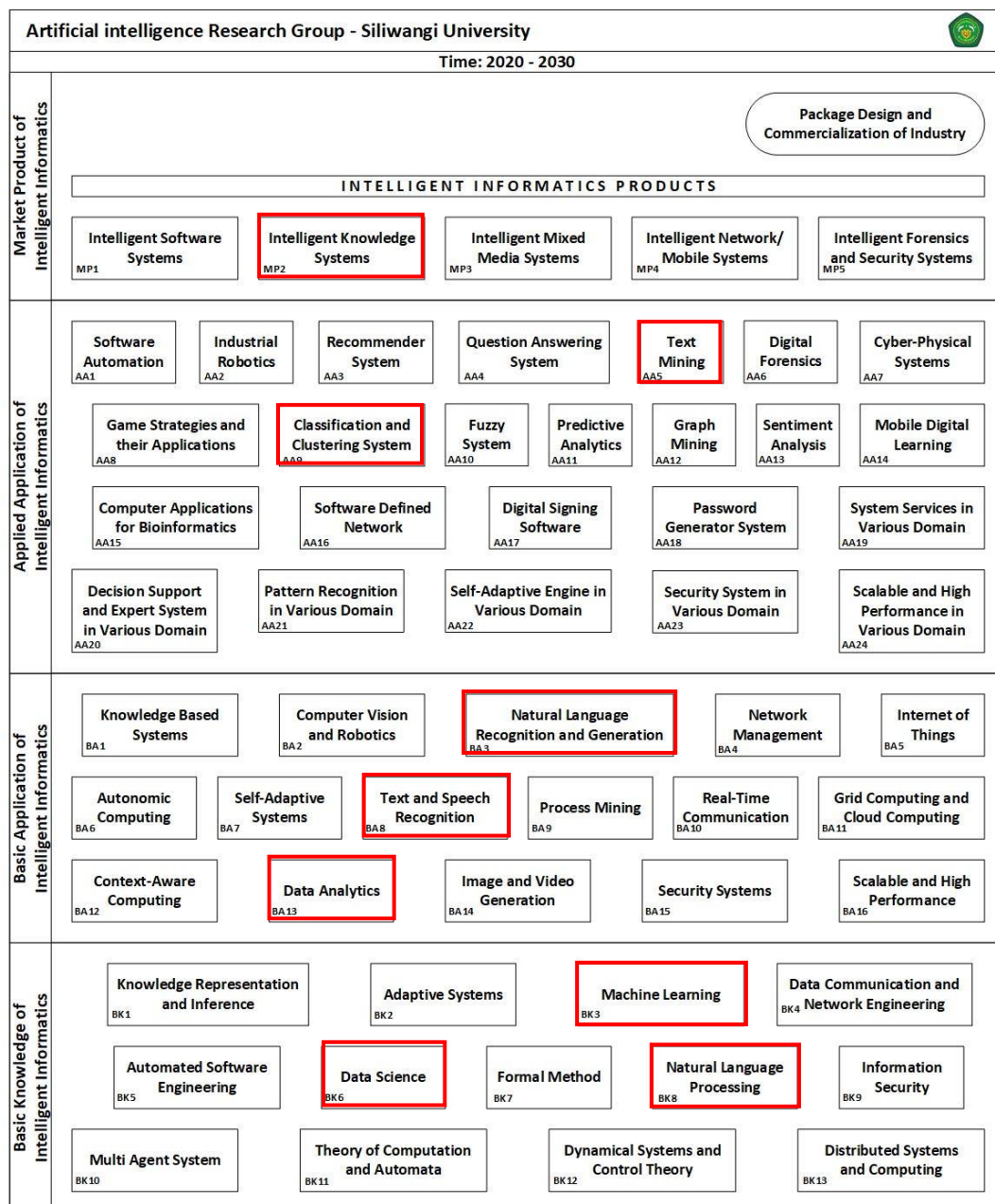
Penelitian ini menggunakan objek teks sarkasme berbahasa Indonesia. Proses penelitian meliputi tahap *preprocessing* data, optimasi *hyperparameter* (Hyperband, Bayesian Optimization, dan BOHB), serta pemodelan menggunakan *pre-train* model IndoBERT. Terdapat empat model yang dikembangkan, diantaranya, tiga model dengan penerapan teknik optimasi *hyperparameter* otomatis dan satu model dikembangkan dengan menggunakan konfigurasi *hyperparameter* dari penelitian sebelumnya. Setiap model yang dikembangkan akan dianalisis dan dibandingkan untuk mengevaluasi pengaruh penerapan optimasi *hyperparameter* terhadap peningkatan kinerja model IndoBERT dalam mendeteksi sarkasme berbahasa Indonesia. Tahap *preprocessing* data bertujuan untuk menyiapkan data sebelum memasuki tahap selanjutnya. Setelah itu, Penerapan teknik optimasi *hyperparameter* akan menghasilkan konfigurasi *hyperparameter* yang optimal, yang kemudian digunakan untuk membangun model

dengan konfigurasi terbaik yang diperoleh. Deteksi sarkasme ini dilakukan dengan melatih model menggunakan dataset sarkasme dan non-sarkasme, sehingga dapat memprediksi kelas kalimat pada data baru.

3.2 Peta Jalan (*Road Map*) Penelitian

Penelitian yang diusulkan dalam laporan 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 terkait *Natural Language Processing, Data Science, Machine Learning, Data Analytics, Text and Speech Recognition, Natural Language Recognition and Generation, Classification and Clustering System, Text Mining* dengan tujuan menghasilkan prototipe *Intelligent Knowledge System* tentang deteksi sarkasme berbahasa Indonesia. Penelitian ini berfokus pada deteksi sarkasme berbahasa Indonesia dengan pemilihan nilai *hyperparameter* optimal secara otomatis. Teknik optimasi *hyperparameter* yang diterapkan dalam penelitian ini meliputi Bayesian Optimization, Hyperband, dan BOHB (Bayesian Optimization Hyperband). Kinerja setiap model hasil optimasi kemudian dibandingkan dan dianalisis terhadap model *baseline* yang ditraining menggunakan konfigurasi *hyperparameter* berdasarkan penelitian sebelumnya. Tujuan dari analisis ini adalah untuk mengevaluasi pengaruh penerapan teknik optimasi terhadap peningkatan kinerja model IndoBERT dalam mendeteksi sarkasme pada

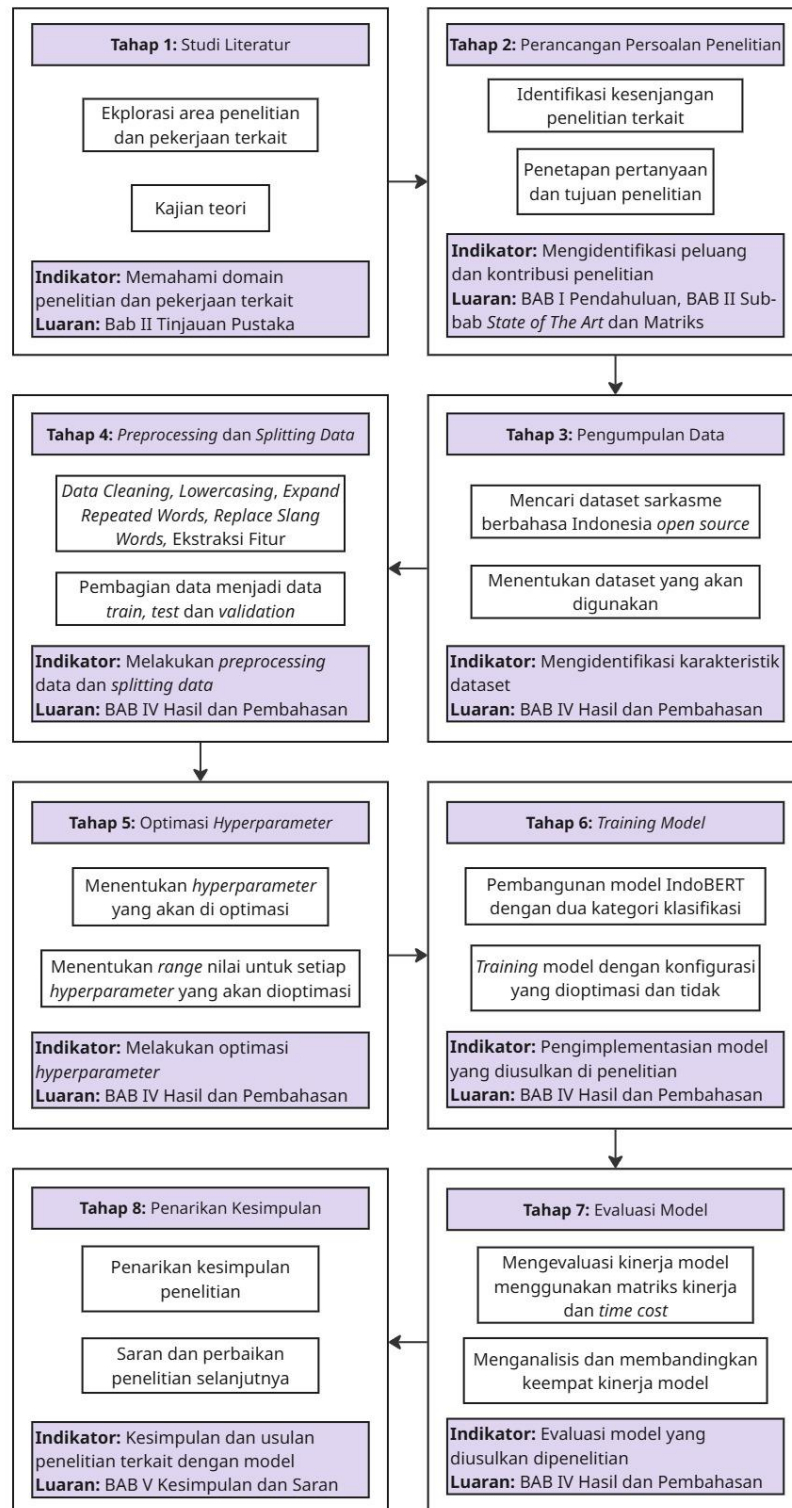
teks berbahasa Indonesia, serta mengidentifikasi teknik optimasi yang menghasilkan kinerja terbaik. Selain itu, penelitian ini merupakan pengembangan dari penelitian skripsi berjudul "Deteksi Sarkasme Menggunakan *Bidirectional Encoder Representations from Transformers* (BERT) pada Teks Bahasa Indonesia" yang disusun oleh (Azka Fauzi Al-Parisi, 2024).



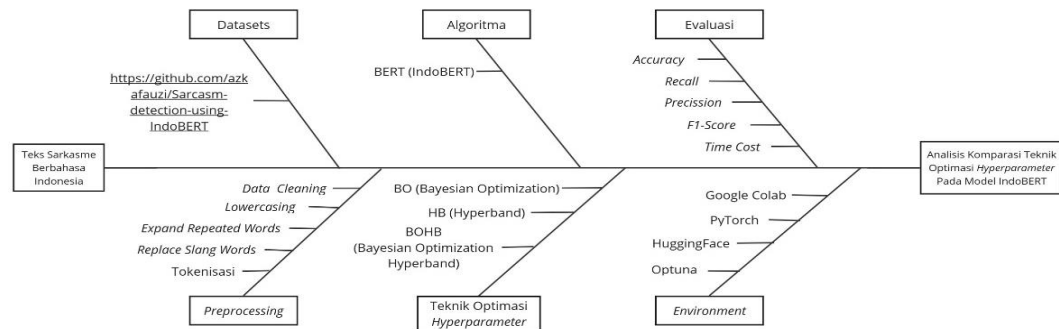
Gambar 3. 1 Roadmap Penelitian (AIS Universitas Siliwangi, 2020)

3.3 Tahapan Penelitian

Penelitian ini bertujuan untuk menghadirkan kebaruan dalam pemilihan nilai *hyperparameter* yang optimal secara otomatis, sehingga menghasilkan konfigurasi terbaik untuk model IndoBERT dalam mendeteksi sarkasme berbahasa Indonesia. Pendekatan ini diwujudkan melalui penerapan teknik optimasi *hyperparameter* untuk menentukan nilai *hyperparameter* optimal pada model IndoBERT. Teknik optimasi *hyperparameter* yang digunakan dalam penelitian ini meliputi Bayesian Optimization, Hyperband, dan BOHB (Bayesian Optimization Hyperband). Diharapkan penerapan teknik optimasi *hyperparameter* ini tidak hanya meningkatkan kinerja model IndoBERT secara keseluruhan dalam mendeteksi sarkasme, tetapi juga meningkatkan efisiensi proses pemilihan nilai *hyperparameter* serta penggunaan sumber daya komputasi yang lebih efisien. Tahapan penelitian yang dilakukan meliputi studi literatur, perancangan persoalan penelitian, *preprocessing* dan *splitting* data, pengumpulan data, optimasi *hyperparameter*, *training* model, penarikan kesimpulan, dan evaluasi model. Tahapan penelitian keseluruhan secara lebih rinci dijelaskan pada Gambar 3.2. Sementara itu, tujuan penelitian ini direpresentasikan dalam diagram *fishbone*. Diagram *fishbone* adalah diagram yang menunjukkan penyebab-penyebab dari sebuah fenomena yang spesifik (Project Management Institute, 2015), yang ditunjukkan pada Gambar 3.3.



Gambar 3.2 Tahapan Penelitian



Gambar 3.3 Diagram *Fishbone*

Bagian ekor diagram *fishbone* pada Gambar 3.3 mewakili objek penelitian, yaitu teks sarkasme berbahasa Indonesia, sedangkan bagian kepala menunjukkan tujuan penelitian, yaitu analisis komparasi teknik optimasi *hyperparameter* pada model IndoBERT. Sementara bagian tulang tengahnya menjelaskan elemen-elemen yang mendukung tercapainya tujuan dari penelitian, dengan keterangan sebagai berikut:

- Dataset*, yaitu data yang akan digunakan sebagai objek dari penelitian berupa dataset teks sarkasme berbahasa Indonesia, diambil dari dataset penelitian serupa sebelumnya yang tersedia pada link *repository* github <https://github.com/azkafauzi/Sarcasm-detection-using-IndoBERT>.
- Preprocessing*, yaitu tahapan yang diterapkan pada dataset yang telah diperoleh, meliputi *data cleaning*, *lowercasing*, *expand repeated words*, *replace slang words*, tokenisasi menggunakan IndoBERT *tokenizer*.
- Algoritma, algoritma atau model yang digunakan dalam penelitian ini untuk mendeteksi sarkasme berbahasa, yaitu *pre-trained* model IndoBERT.
- Teknik Optimasi *Hyperparameter*, yaitu teknik yang digunakan untuk mencari konfigurasi *hyperparameter* yang optimal secara otomatis. Teknik

optimasi *hyperparameter* yang digunakan pada penelitian ini meliputi Bayesian Optimization, Hyperband, dan BOHB (Bayesian Optimization Hyperband)

- e. Evaluasi, evaluasi dilakukan untuk mengukur kinerja model dengan mengevaluasi hasil dari matriks kinerja (*accuracy, precision, recall, F1-score*) dan *time cost*.
- f. *Environment*, yaitu lingkungan pengembangan model meliputi Google Colab sebagai Jupyter Notebook *environment* serta menggunakan *pytorch*, *huggingface*, dan *optuna* sebagai *library* utama yang digunakan untuk pengembangan model.

3.3.1 Studi Literatur

Tahap ini bertujuan untuk mengeksplorasi bidang penelitian serta mengidentifikasi penelitian terdahulu yang relevan. Kajian teori dilakukan untuk memahami konsep dasar dan perkembangan terbaru dalam bidang yang diteliti. Luaran dari tahap ini berupa tinjauan pustaka yang disajikan dalam Bab II.

3.3.2 Perancangan Persoalan Penelitian

Tahap ini melakukan identifikasi kesenjangan penelitian berdasarkan literatur yang telah dikaji untuk merumuskan RP (*Research Problems*). Selanjutnya, ditetapkan RQ (*Research Questions*) serta RO (*Research Objectives*) untuk menentukan kontribusi akademik yang dapat dihasilkan. Luaran dari tahap ini meliputi bagian pendahuluan serta subbab *State of The Art* dalam Bab II.

3.3.3 Pengumpulan Data

Tahap ini mencakup pencarian dataset sarkasme berbahasa Indonesia yang bersifat *open source* serta memilih dataset yang akan digunakan. Dataset yang digunakan pada penelitian ini berasal dari penelitian serupa sebelumnya, yang tersedia pada *link repository* Github <https://github.com/azkafauzi/Sarcasm-detection-using-IndoBERT> diakses pada 11 Maret 2025. Setelah itu, dilakukan proses identifikasi untuk memahami karakteristik dataset. Dataset berupa file Excel yang terdiri dari 4.400 baris data dengan dua kolom, yaitu *text* berisi kalimat dan *sarcasm* berisi label. Kolom *sarcasm* memiliki dua kelas, yakni sarkasme dan non-sarkasme, dengan jumlah data yang seimbang di masing-masing kelas. Luaran dari tahap ini didokumentasikan dalam Bab IV.

3.3.4 Preprocessing dan Splitting Data

Dataset yang telah dikumpulkan selanjutnya dilakukan tahap *preprocessing*. Proses ini melibatkan *data cleaning*, *lowercasing*, *expand repeated words*, *replace slang words* dan ekstraksi fitur untuk memastikan kesiapan data dalam model yang akan dikembangkan. Setelah itu, dilakukan juga *splitting data* menjadi 3 bagian, yaitu dataset *training*, *testing* dan *validation* dengan rasio 60%:20%:20%. Luaran dari tahap ini dijelaskan dalam Bab IV.

3.3.5 Optimasi Hyperparameter

Optimasi *hyperparameter* dilakukan dengan menentukan teknik optimasi *hyperparameter* yang akan digunakan. Teknik yang digunakan meliputi Hyperband, Bayesian Optimization, BOHB (Bayesian Optimization Hyperband). Selanjutnya, ditentukan *hyperparameter* yang akan dioptimasi beserta rentang

pencarian nilainya. Luaran dari tahapan ini berupa konfigurasi *hyperparameter* yang optimal yang akan digunakan pada tahap selanjutnya. Hasil dari tahapan ini didokumentasikan dalam BAB IV.

3.3.6 Training Model

Training model dilakukan sebanyak empat kali, di mana tiga di antaranya menggunakan konfigurasi *hyperparameter* hasil optimasi, sedangkan satu lainnya menggunakan konfigurasi dari penelitian sebelumnya sebagai pembandingan. Model yang dikembangkan merupakan IndoBERT dengan dua kategori klasifikasi. Hasil dari tahapan ini didokumentasikan dalam BAB IV.

3.3.7 Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja model berdasarkan metrik kinerja dan *time cost*. Analisis dilakukan dengan membandingkan kinerja keempat model yang telah dilatih guna menilai efektivitas penerapan teknik optimasi *hyperparameter*. Luaran dari tahap ini didokumentasikan dalam Bab IV.

3.3.8 Penarikan Kesimpulan

Tahap ini membuat kesimpulan dirumuskan berdasarkan hasil evaluasi model. Selain itu, disusun rekomendasi serta perbaikan untuk penelitian selanjutnya. Luaran dari tahap ini disajikan dalam Bab V yang berisi kesimpulan dan saran.