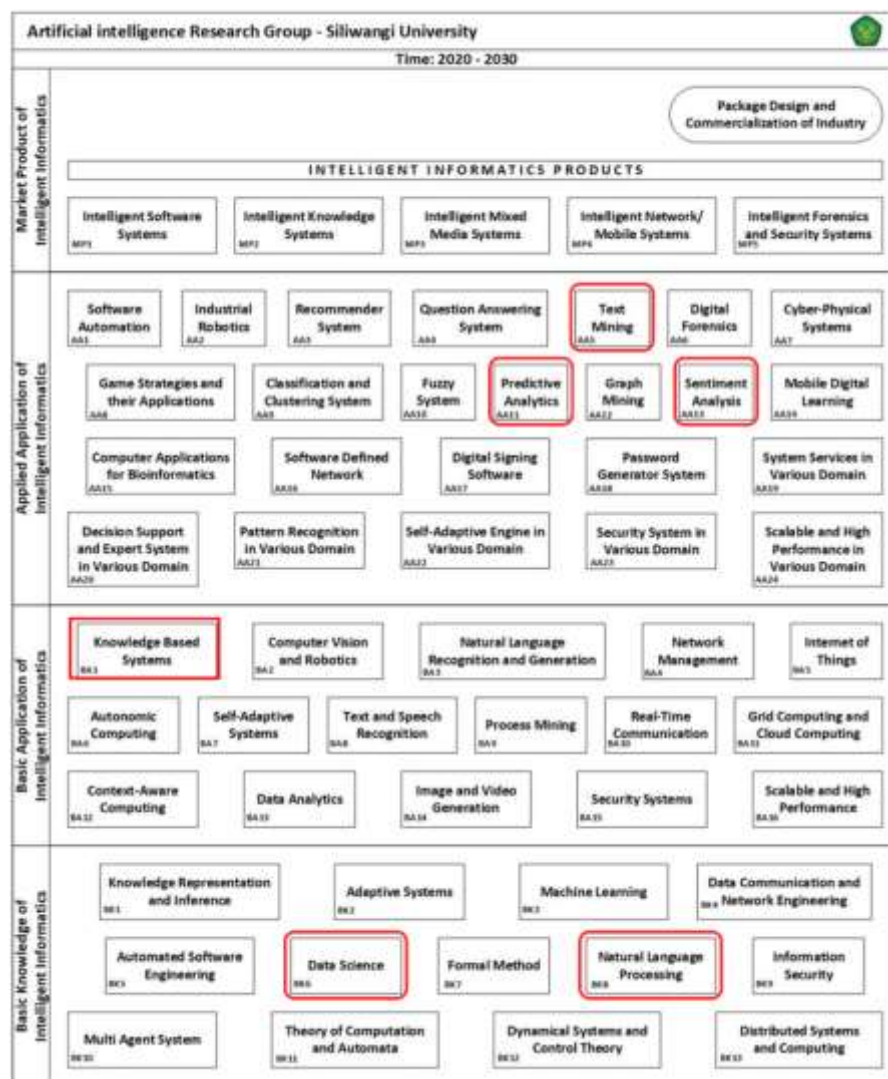


BAB III

METODOLOGI PENELITIAN

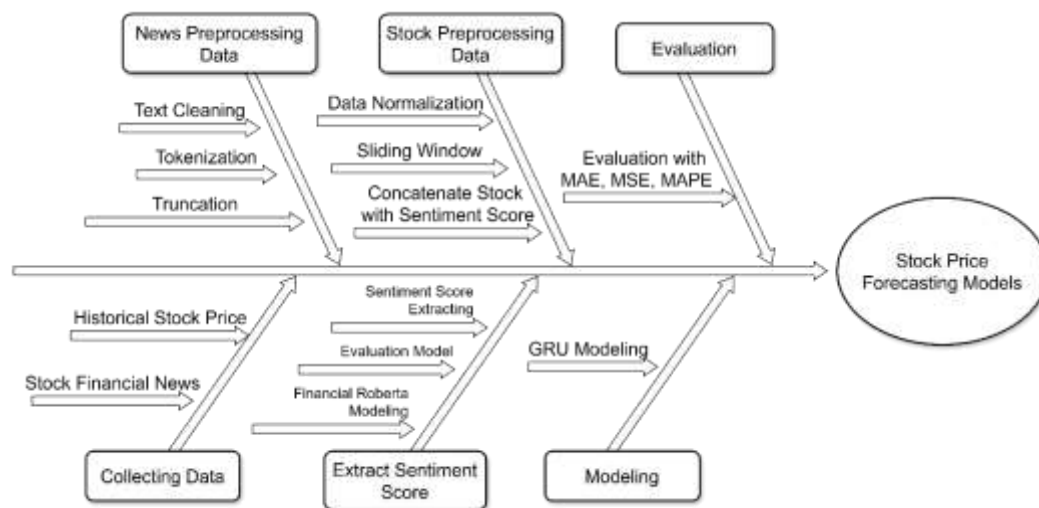
3.1 Roadmap Penelitian

Rencana penelitian ini secara keseluruhan selaras dengan *roadmap* Kelompok Keahlian Informatika dan Sistem Inteligen (ISI). Berikut merupakan *roadmap* penelitian seperti yang telah digambarkan pada Gambar 3.1.



Gambar 3.1 *Roadmap* Penelitian

Pada gambar 3.1 dapat dilihat bahwa *roadmap* dapat dilihat bahwa peta jalan penelitian ini dimulai dengan pemahaman konsep dasar dalam bidang informasi cerdas, khususnya terkait dengan *forecasting* saham yang menggabungkan analisis sentimen dan data historis saham. Langkah pertama adalah memperoleh pemahaman mendalam tentang sistem berbasis pengetahuan (*knowledge-based system*) yang memanfaatkan data historis dan skor sentimen sebagai landasan utama dalam proses prediksi. Berikut adalah gambaran proses yang digunakan dalam sistem pada Gambar 3.2.



Gambar 3.2 Fishbone Penelitian

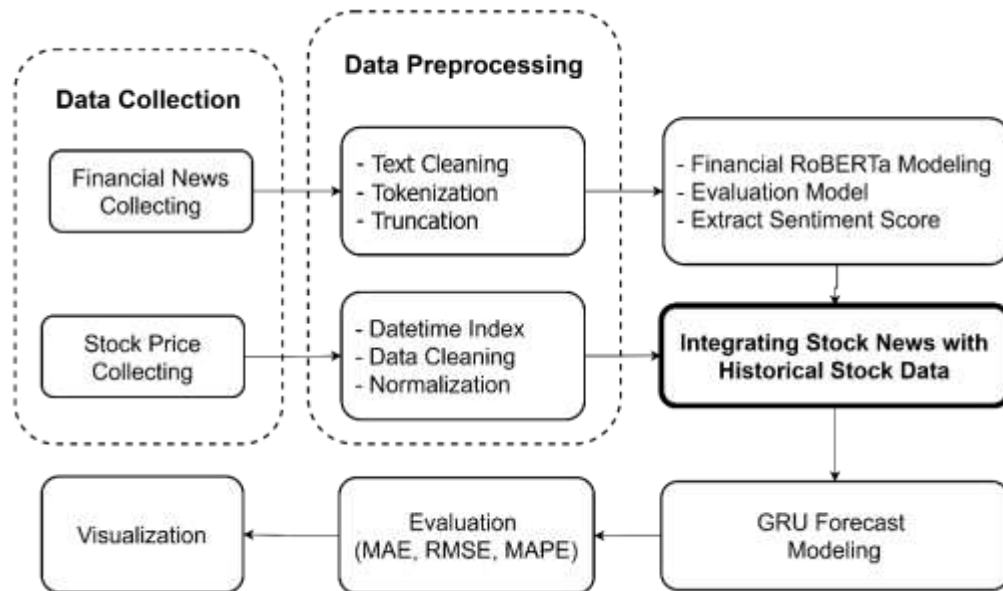
Gambar 3.2. menggambarkan *Knowledge Based System* Penelitian yaitu integrasi antara analisis sentimen dengan historis saham untuk memprediksi harga saham. Proses penelitian diawali dengan pengumpulan data yang mencakup berita saham dan data historis saham. Selanjutnya, berita yang diperoleh diproses dalam tahap preprocessing, yang mencakup *Text Cleaning*, *Tokenization*, *Padding* dan *Truncation*. agar teks berita sesuai dengan kebutuhan model RoBERTa, yang memerlukan panjang tensor yang seragam untuk pemrosesan yang optimal.

Setelah tahap *preprocessing* untuk data berita, dilakukan modeling menggunakan *Financial RoBERTa* untuk menganalisis sentimen setiap berita yang ada lalu dilakukan evaluasi terhadap model yang telah dibuat, setelah akurasi model sudah cukup baik dilanjutkan dengan ekstraksi skor sentimen untuk mengukur sentimen positif, negatif, atau netral dari teks berita setiap saham. Hasil sentimen ini kemudian dikombinasikan dengan data historis saham seperti harga pembukaan, penutupan dan *volume* perdagangan yang telah di *preprocessing (Data Cleaning dan Normalization)*. Setelah itu, dilakukan *concatenating data*, di mana sentimen skor dan historis saham digabungkan menjadi satu dataset dengan indeks tanggal yang sama untuk digunakan dalam pemodelan. Model *forecasting* berbasis *GRU* digunakan untuk memprediksi pergerakan harga saham berdasarkan pola historis dan pengaruh sentimen berita. Terakhir, model dievaluasi menggunakan matrik seperti *RMSE*, *MAE* dan *MAPE* untuk memastikan keakuratan dan keandalannya dalam mendukung pengambilan keputusan investasi.

3.2 Metode Penelitian

Capaian dalam penelitian ini diharapkan dapat menghasilkan model peramalan harga saham yang lebih akurat dengan mengintegrasikan analisis sentimen berita saham dan data historis saham. Pendekatan ini memungkinkan pemodelan yang lebih informatif dengan mempertimbangkan sentimen pasar terhadap pergerakan harga saham. Untuk mencapai tujuan tersebut, penelitian ini dilakukan melalui dua tahap utama, yaitu analisis sentimen dari berita saham dan pemrosesan data historis saham, sebelum akhirnya digabungkan untuk pemodelan

prediksi harga saham. Tahapan penelitian yang akan dilakukan dapat dilihat pada Gambar 3.3.



Gambar 3.3 Tahapan Penelitian

A. Data Collection

Pengumpulan data merupakan tahap awal yang penting dalam penelitian ini. Data yang digunakan terdiri dari berita finansial untuk analisis sentimen serta data historis saham sebagai dasar peramalan harga saham.

1. Sentimen Analisis

Data yang digunakan dalam penelitian ini terdiri dari dua jenis utama, yaitu dataset *Indonesian Financial PhraseBank* dan berita finansial dari sumber berita daring. *Indonesian Financial PhraseBank* merupakan kumpulan frasa keuangan berbahasa Indonesia yang tersedia di Hugging Face (intanm/indonesian-Financial-phrasebank). Dataset ini digunakan untuk melakukan *fine-tuning* pada model

Financial RoBERTa agar dapat mengenali sentimen dalam konteks keuangan dengan lebih akurat.

Data berita finansial dikumpulkan dari dua sumber utama, yaitu Bisnis.com dan CNBCIndonesia. Berita yang dikumpulkan mencakup informasi terkait saham tiga bank besar di Indonesia, yaitu BBRI, BBKA, dan BMRI. Data ini akan digunakan sebagai input dalam model analisis sentimen guna mengklasifikasikan sentimen berita terkait pergerakan saham dari ketiga bank tersebut.

2. Peramalan Saham

Data historis saham diperoleh dari situs Investing.com yang mencakup berbagai fitur seperti harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), dan *volume* perdagangan. Dalam penelitian ini, harga penutupan (*close*) digunakan sebagai target utama dalam peramalan saham.

B. Data Preprocessing

Tahapan *preprocessing* bertujuan untuk mengolah data agar siap digunakan dalam pemodelan.

1. Sentimen Analisis

Sebelum dianalisis menggunakan model *Financial RoBERTa*, data berita finansial terlebih dahulu melalui beberapa tahap preprocessing teks, yaitu:

- a. *Text Cleaning*: Membersihkan teks dari karakter khusus, tanda baca yang tidak relevan, serta menghapus tautan dan angka yang tidak diperlukan.

- b. *Tokenization*: Memecah teks menjadi unit yang lebih kecil (*tokens*) menggunakan metode *subword tokenization* agar dapat dipahami oleh model.
- c. *Truncation*: Menyamakan panjang tensor dengan memangkas token pada input yang panjangnya melebihi batas maksimum tensor model.

2. Peramalan Saham

Data historis saham juga perlu diproses sebelum digunakan dalam pemodelan. Tahapan preprocessing untuk data saham meliputi:

- a. *Datetime Index*: Mengubah indeks dataset menggunakan Tanggal dengan format *datetime*.
- b. *Normalization*: Melakukan normalisasi nilai pada dataset agar skala data seragam, yang bertujuan untuk meningkatkan stabilitas dan kinerja model prediksi.
- c. *Sliding Window*: Membentuk data historis ke dalam *window* berukuran tetap, di mana setiap jendela memuat sejumlah titik data sebelumnya yang digunakan untuk memperkirakan nilai pada titik waktu berikutnya.

C. Modeling

Dilakukan pemodelan untuk analisis sentimen terlebih dahulu, karena hasil dari hasil dari analisis sentimen yaitu sentimen skor setiap berita dibutuhkan untuk pemodelan pada peramalan saham.

1. Sentimen Analisis

Model *Financial RoBERTa* digunakan untuk melakukan analisis sentimen terhadap berita finansial yang telah diproses sebelumnya. Proses ini mencakup:

- a. *Training* dan *Fine-Tuning* model agar dapat mengenali pola sentimen dari berita finansial.
- b. Evaluasi Model untuk mengukur performa model dalam mengklasifikasikan sentimen dengan akurat.
- c. Ekstraksi Skor Sentimen dari berita saham, yang nantinya akan dikombinasikan dengan data historis saham dalam tahap peramalan.

2. Peramalan Saham

Peramalan saham menggunakan pendekatan *deep learning* dengan *Gated Recurrent Unit (GRU)*. Proses yang dilakukan dalam tahap ini meliputi:

- a. Menggabungkan Skor Sentimen dari berita dengan Data Historis Saham dengan menggunakan tanggal yang sama sebagai index.
- b. Pelatihan Model *GRU* menggunakan dataset yang telah dikombinasikan untuk mempelajari pola pergerakan harga saham berdasarkan histori dan sentimen pasar.

D. Evaluation

Tahapan evaluasi dilakukan untuk mengukur performa model dalam melakukan prediksi harga saham. Evaluasi model dilakukan dengan menggunakan beberapa matrik berikut:

- a. *Mean Absolute Error (MAE)*: Mengukur rata - rata dari nilai absolut selisih antara nilai aktual dan nilai prediksi.
- b. *Root Mean Squared Error (RMSE)*: Mengukur akar dari rata - rata kuadrat selisih antara nilai aktual dan prediksi.

- c. *Mean Absolute Percentage Error (MAPE)*: Mengukur persentase kesalahan prediksi terhadap nilai aktual untuk memahami tingkat akurasi model.

E. *Visualization*

Tahap akhir dari penelitian ini diwujudkan dalam bentuk visualisasi *dashboard* interaktif menggunakan *Looker Studio*, yang dirancang untuk menginterpretasikan dan membandingkan hasil dari beberapa skenario model peramalan harga saham. Model yang dibandingkan mencakup model berbasis data historis saja, serta model yang mengintegrasikan data historis dengan skor sentimen dari berita keuangan yang diperoleh dari situs CNBCIndonesia dan Bisnis.com. Dashboard ini menampilkan perbandingan akurasi antara kedua pendekatan dengan menggunakan matrik evaluasi seperti RMSE, RMSE dan MAPE, serta visualisasi garis antara harga aktual dan hasil prediksi dari masing-masing model.

Dashboard ini juga menyajikan analisis pengaruh sentimen berita terhadap pergerakan harga saham, dengan mengelompokkan berita ke dalam kategori faktor internal maupun eksternal yang memengaruhi harga saham. Dengan pendekatan ini, investor dapat melihat keterkaitan antara jenis sentimen dan isu fundamental tertentu dengan arah dan besarnya perubahan harga saham yang terjadi setelah berita dirilis. Visualisasi ini diharapkan dapat memberikan wawasan yang lebih menyeluruh mengenai pengaruh integrasi data sentimen terhadap peningkatan akurasi peramalan serta hubungan antara berita keuangan dan dinamika harga saham perbankan.