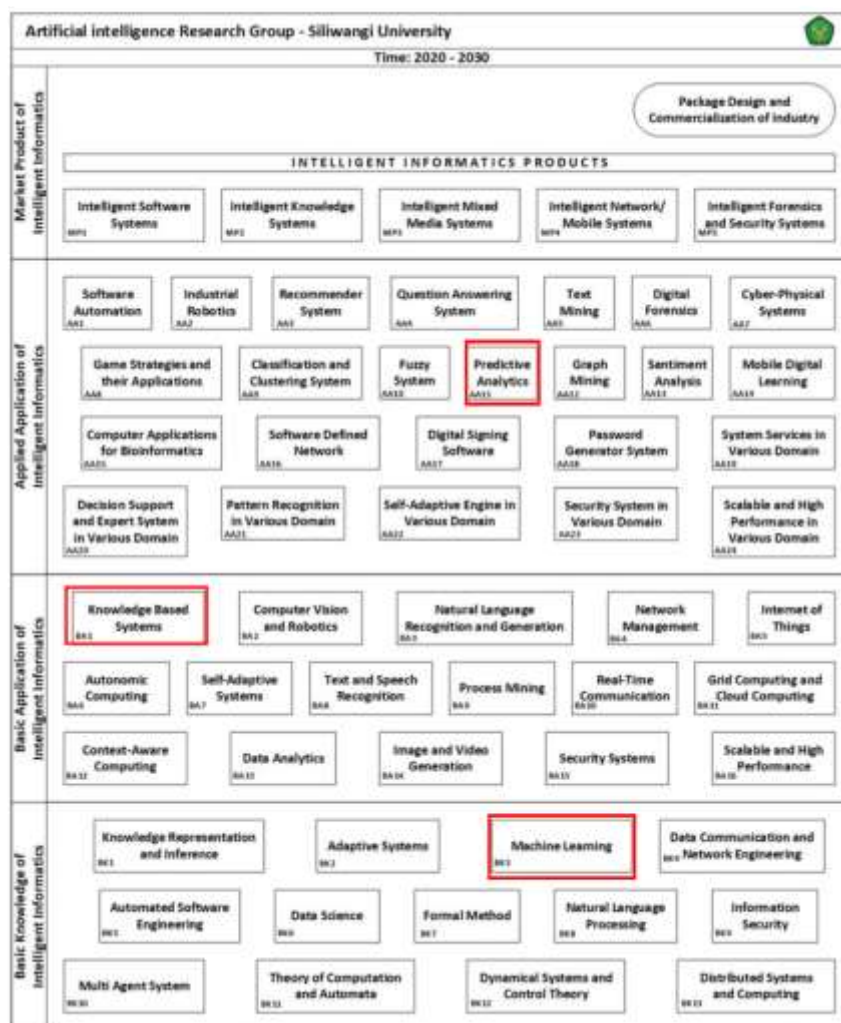


BAB III

METODE PENELITIAN

3.1 Peta Jalan (Road Map) Penelitian

Topik penelitian yang diangkat dalam penelitian ini sejalan dengan Peta Jalan Kelompok Keahlian Informatika dan Sistem Inteligen (ISI). Berikut merupakan peta jalan penelitian seperti yang telah disajikan pada Gambar 3.1

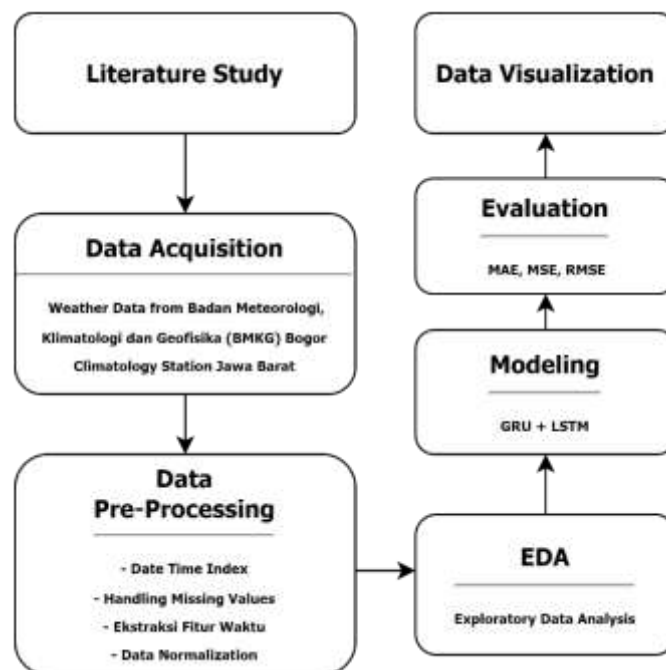


Gambar 3. 1 Peta Jalan Penelitian (AIS Universitas Siliwangi, 2024)

Penelitian ini menggabungkan tiga disiplin ilmu yakni *predictive analysis*, *knowledge-based system*, dan *machine learning*. *Predictive analysis* digunakan untuk mengidentifikasi pola dalam data historis guna memperkirakan tren cuaca. *Knowledge-based system* berperan dalam membangun pemodelan sebagai dasar pengetahuan. Sementara itu, *machine learning* diterapkan melalui model LSTM dan GRU untuk peramalan cuaca.

3.2 Tahapan Penelitian

Penelitian ini melibatkan serangkaian langkah atau prosedur yang sistematis. Gambar 3.2 menyajikan alur yang menggambarkan tahapan penelitian secara keseluruhan secara lebih terstruktur dan komprehensif.



Gambar 3. 2 Tahapan Penelitian

3.2.1. *Literature study*

Literature study dalam penelitian ini merujuk pada artikel-artikel ilmiah nasional dan internasional, serta sumber lain yang berkaitan dengan peramalan cuaca, model LSTM, dan model GRU. Kajian literatur ini dilakukan untuk memperoleh pemahaman yang komprehensif mengenai peramalan cuaca berbasis *deep learning*, termasuk keunggulan dan keterbatasan masing-masing model. Selain itu, *literature study* juga digunakan sebagai landasan dalam merumuskan pendekatan yang tepat dalam penelitian ini, dan mengidentifikasi *research gap*.

3.2.2. *Data Acquisition*

Data yang digunakan dalam penelitian ini adalah data cuaca harian dari Stasiun Klimatologi Jawa Barat Bogor yang bersumber dari situs resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dengan rentang waktu dari 1 Juli 2020 sampai 7 Juli 2024 sebanyak 1467 baris.

Data ini mencakup beberapa variabel cuaca yang akan digunakan sebagai fitur dalam model peramalan. Penelitian ini menggunakan pendekatan *multioutput*, di mana model akan memprediksi lebih dari satu variabel target. Variabel target yang dipilih mencakup suhu maksimum, suhu minimum, curah hujan, dan kecepatan angin yang mana merupakan variabel-variabel yang memengaruhi potensi terjadinya cuaca ekstrem.

3.2.3. *Data Pre-processing*

Kualitas data yang digunakan dalam penelitian sangat berpengaruh terhadap kinerja model. Oleh karena itu, dilakukan serangkaian tahapan *preprocessing* untuk

memastikan data yang digunakan bersih, terstruktur, dan siap untuk diproses oleh model. Tahapan *preprocessing* yang diterapkan pada penelitian ini terdiri dari beberapa langkah yakni:

1. *Date Time Index*

Tahapan *preprocessing* pertama yang dilakukan adalah mengkonversi kolom ‘tanggal’ pada dataset ke dalam format *date time Index* untuk memastikan keteraturan data berdasarkan urutan waktu. Konversi ini memungkinkan berbagai operasi berbasis waktu, seperti resampling, interpolasi, dan pengelompokan data, sehingga analisis dan pemrosesan data dapat dilakukan dengan lebih efisien. Selain itu, penggunaan *date time index* membantu model dalam mengenali pola temporal dengan lebih baik.

2. *Handling Missing Values*

Tahapan *preprocessing* yang dilakukan selanjutnya adalah *handling missing values* untuk menjaga kualitas data sebelum digunakan dalam pelatihan model. Mengingat data cuaca memiliki ketergantungan temporal, metode *forward fill* digunakan dengan mengisi *missing values* menggunakan data sebelumnya dalam data *time series*. Pendekatan ini dipilih untuk mempertahankan kontinuitas data tanpa mengubah pola distribusi yang dapat memengaruhi hasil prediksi.

3. Ekstraksi Fitur Waktu

Ekstraksi fitur waktu dari kolom tanggal turut dilakukan guna memperkaya informasi temporal yang digunakan dalam proses pelatihan model. Fitur waktu yang diambil meliputi tahun (*year*), bulan (*month*), dan hari (*day*). Ekstraksi ini dilakukan karena variabel-variabel cuaca umumnya dipengaruhi oleh pola musiman

dan siklus waktu tertentu. Dengan menambahkan komponen waktu sebagai fitur tambahan, model diharapkan dapat mengenali pola berulang dan variasi musiman yang terdapat dalam data cuaca.

4. *Data Normalization*

Tahapan terakhir *preprocessing* dalam penelitian ini adalah melakukan normalisasi data untuk memastikan setiap fitur berada dalam skala yang seragam, sehingga mempercepat konvergensi model. Metode normalisasi yang digunakan dalam penelitian ini adalah *Min-Max Scaling* agar nilai setiap fitur berada dalam rentang $[0,1]$ tanpa mengubah distribusi relatif data. Proses ini penting dalam model *deep learning* seperti *hybrid GRU-LSTM*, karena sensitivitas terhadap perbedaan skala antar fitur dapat memengaruhi kualitas pembelajaran dan akurasi prediksi.

3.2.1. *Exploratory Data Analysis (EDA)*

Tahap *Exploratory Data Analysis (EDA)* dilakukan dalam penelitian ini setelah data dikumpulkan dan sebelum masuk ke tahap pemodelan. Tujuan utama dari EDA adalah untuk memperoleh pemahaman awal terhadap struktur dan karakteristik data digunakan. Proses EDA mencakup analisis distribusi masing-masing variabel target yakni T_n , T_x , RR , ff_x , identifikasi pola musiman berdasarkan bulan dan kuartal, serta analisis hubungan antar variabel menggunakan korelasi.

Visualisasi yang digunakan dalam tahap ini antara lain grafik boxplot untuk melihat sebaran nilai berdasarkan waktu, plot garis untuk menganalisis tren data, dan heatmap korelasi untuk mengevaluasi kekuatan hubungan antar fitur numerik.

3.2.2. *Modeling*

Proses *modeling* dilakukan dengan membangun arsitektur *hybrid* GRU-LSTM untuk peramalan cuaca. Arsitektur model akan dirancang dengan mengombinasikan lapisan GRU dan LSTM, di mana jumlah lapisan dan unit dalam setiap lapisan akan ditentukan guna memperoleh struktur optimal. Model ini akan dilatih menggunakan dataset cuaca yang telah diproses, dengan menggunakan variabel-variabel yang memengaruhi potensi cuaca ekstrem.

3.2.3. *Evaluation*

Tahap *evaluation* dilakukan untuk menilai kemampuan model. Evaluasi model dalam penelitian ini akan menggunakan beberapa metrik evaluasi yakni:

1. MAE (*Mean Absolute Error*): mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual
2. MSE (*Mean Squared Error*): menghitung rata-rata kuadrat selisih antara nilai yang diprediksi dan nilai aktual
3. RMSE (*Root Mean Squared Error*): akar kuadrat dari MSE dan memberikan ukuran kesalahan dalam satuan yang sama dengan data asli

3.2.4. *Data Visualization*

Data Visualization merupakan tahap akhir dalam penelitian ini yang bertujuan untuk menganalisis dan menginterpretasikan hasil prediksi model. Visualisasi dilakukan terhadap data *output* model guna memahami pola, tren, serta

karakteristik dari variabel. Tahap ini penting untuk menilai sejauh mana model mampu merepresentasikan kondisi cuaca yang sebenarnya.

Salah satu aspek utama dalam tahap ini adalah mengidentifikasi indikasi cuaca ekstrem. Data hasil peramalan dikategorikan berdasarkan ambang batas variabel cuaca yang diperoleh dari data historis dan rata-rata kondisi cuaca (Weather Spark, n.d.). Apabila hasil peramalan melebihi atau berada di luar rentang normal, maka kondisi tersebut dapat dikategorikan sebagai potensi cuaca ekstrem. Ambang batas variabel cuaca yang digunakan dapat dilihat pada Tabel 3.1

Tabel 3. 1 Ambang batas variabel cuaca ekstrem Bogor (Weather Spark, n.d.)

No	Bulan	Suhu Maks	Suhu Min	Curah Hujan	Kecepatan Angin
1	Januari	29°C	22°C	275,2mm	10.1
2	Februari	29°C	22°C	256,0mm	9.8
3	Maret	30°C	22°C	186,5mm	8.1
4	April	31°C	22°C	162,4mm	6.8
5	Mei	31°C	22°C	115,4mm	6.3
6	Juni	31°C	21°C	70,2mm	6.4
7	Juli	31°C	21°C	58,8mm	6.5
8	Agustus	31°C	21°C	53,2mm	6.6
9	September	32°C	21°C	66,9mm	6.4
10	Oktober	32°C	22°C	110,3mm	6.3
11	November	31°C	22°C	169,8mm	7.1
12	Desember	30°C	22°C	198,9mm	9.2