

BAB I PENDAHULUAN

1.1 Latar belakang

Prediksi deret waktu telah menjadi aspek krusial dalam berbagai bidang, termasuk keuangan, energi, kesehatan, dan transportasi (Liu et al., 2020; Rodrigues et al., 2019; Sen & Datta Chaudhuri, 2021). Kemampuan untuk mengenali pola temporal dalam data historis dan menghasilkan prediksi yang akurat memiliki dampak signifikan terhadap pengambilan keputusan strategis, efisiensi operasional, dan penghematan biaya. Di Kota Makassar, data cuaca menjadi salah satu informasi penting yang dapat digunakan untuk perencanaan aktivitas masyarakat, mitigasi bencana, dan pengelolaan sumber daya.

Namun, tantangan dalam deret waktu sering kali mencakup data yang tidak lengkap, hubungan antar variabel yang kompleks, dan ketergantungan temporal yang panjang (Bao & Lu, 2024). Model prediktif tradisional seperti ARIMA (Jere et al., 2017) dan VAR (Sims, 1980), meskipun telah lama digunakan, sering kali kurang efektif dalam menangani data multivariat yang kompleks dan pola temporal yang tidak linier (Yang & Lu, 2022). Dalam beberapa tahun terakhir, pendekatan berbasis *deep learning* seperti Long Short-Term Memory (LSTM) dan Gated Recurrent Units (GRU) telah menunjukkan peningkatan performa, tetapi sering kali masih memiliki keterbatasan dalam hal interpretabilitas dan efisiensi komputasi untuk prediksi *multi-horizon* (Lim et al., 2021).

Sebagai solusi yang menjanjikan, Temporal Fusion Transformer (TFT) dikembangkan sebagai model yang mampu menggabungkan performa prediksi tinggi dengan interpretabilitas yang mendalam. TFT menggunakan mekanisme *self-attention* untuk menangkap hubungan temporal jangka panjang, serta lapisan *gating* untuk menyaring informasi yang relevan. Model ini juga dapat memanfaatkan data statis dan dinamis secara bersamaan, sehingga cocok untuk skenario deret waktu yang kompleks (Lim et al., 2021). Selain itu, model TFT menunjukkan performa yang lebih baik dalam peramalan jangka pendek dibandingkan dengan lima model berbasis *transformer* lainnya serta model klasik (Joseph et al., 2024).

Selain pengembangan model, visualisasi hasil prediksi juga menjadi aspek penting untuk mendukung interpretasi data dan pengambilan keputusan (Suresh & P, 2022). Dalam hal ini, dasbor interaktif menjadi solusi efektif untuk menyajikan wawasan prediktif secara dinamis, memfasilitasi pengguna dalam eksplorasi data dan interpretasi hasil prediksi (Cohen et al., 2021).

Dengan mempertimbangkan tantangan dan peluang dalam prediksi deret waktu, penelitian ini bertujuan untuk mengevaluasi performa algoritma Temporal Fusion Transformer (TFT) dalam memprediksi curah hujan di Kota Makassar. Selain itu, penelitian ini juga berfokus pada pembangunan dasbor interaktif yang dapat menyajikan hasil prediksi secara visual dan intuitif. Dasbor ini diharapkan mampu memberikan wawasan yang lebih mendalam terkait pola curah hujan serta mendukung pengambilan keputusan berbasis data dalam berbagai sektor, seperti perencanaan kota, mitigasi bencana, dan pengelolaan sumber daya air.

1.2 Rumusan Masalah

1. Bagaimana performa algoritma Temporal Fusion Transformer (TFT) dalam memprediksi curah hujan Kota Makassar?
2. Bagaimana menyajikan hasil prediksi model Temporal Fusion Transformer (TFT) dan analisis dasar data cuaca Makassar melalui dasbor interaktif?

1.3 Tujuan dan Manfaat

Tujuan penelitian difokuskan pada pembangunan dan pengukuran performa model prediksi curah hujan berbasis Temporal Fusion Transformer (TFT) serta implementasi dasbor interaktif untuk menyajikan hasil prediksi dan analisis data cuaca.

Penelitian ini memiliki beberapa tujuan utama yang ingin dicapai, yaitu:

1. Membangun model prediksi curah hujan menggunakan algoritma Temporal Fusion Transformer (TFT) dengan data cuaca Kota Makassar dan mengukur performanya.
2. Membangun dasbor interaktif yang menyajikan hasil prediksi model Temporal Fusion Transformer (TFT) serta analisis dasar data cuaca Kota Makassar.

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan kontribusi dalam pengembangan teknologi prediksi berbasis AI.
2. Mendukung pemerintah daerah, instansi terkait, dan masyarakat dalam pengambilan keputusan terkait cuaca.

1.4 Batasan

Penelitian ini memiliki beberapa keterbatasan, yaitu:

1. *Dataset* yang digunakan masih terbatas pada data historis dan data BMKG, tanpa mempertimbangkan data dari sumber lain.
2. Model TFT dapat terus dioptimalkan untuk meningkatkan akurasi prediksi, terutama untuk kejadian cuaca ekstrem.

1.5 Teori

1.5.1 Data Cuaca dan Tantangannya

Data cuaca merupakan kumpulan informasi yang menggambarkan kondisi atmosfer di suatu lokasi dan waktu tertentu. Jenis-jenis data cuaca meliputi suhu udara, yang mengukur tingkat panas atau dingin udara dan digunakan dalam berbagai analisis perubahan iklim; kelembapan udara, yang menunjukkan kadar uap air di atmosfer dan memengaruhi pembentukan hujan dan kabut; curah hujan, yang merekam jumlah air yang jatuh ke permukaan bumi dan relevan dalam memprediksi potensi banjir serta ketersediaan air; serta tekanan udara, kecepatan dan arah angin, yang sangat penting dalam analisis pola angin dan prediksi badai. Data tambahan seperti radiasi matahari, visibilitas, dan indeks panas juga sering digunakan untuk mendukung aplikasi energi terbarukan, kesehatan, dan mitigasi bencana (M et al., 2024).

Namun, pengolahan data cuaca menghadapi berbagai tantangan. Salah satunya adalah volume data yang sangat besar, karena data cuaca dihasilkan secara terus-menerus dari berbagai sumber seperti satelit, radar, stasiun cuaca, dan model numerik. Proses penyimpanan dan analisis data ini membutuhkan infrastruktur komputasi yang canggih untuk memprosesnya dalam waktu nyata. Selain itu, keragaman format dan resolusi data dari berbagai perangkat sering kali menjadi hambatan dalam integrasi dan analisis data yang efektif (Paul et al., 2024).

Ketidaklengkapan data atau *data gaps* juga sering terjadi akibat kondisi cuaca ekstrem atau gangguan teknis, yang dapat memengaruhi keakuratan analisis dan prediksi. Tantangan lainnya adalah ketidakpastian dalam model cuaca, yang sering kali disebabkan oleh kompleksitas sistem atmosfer yang dinamis dan non-linear. Hal ini semakin sulit ketika mencoba memprediksi cuaca dalam jangka panjang atau di wilayah dengan data observasi yang terbatas (Yao et al., 2022).

Kualitas data yang buruk juga menjadi masalah umum, *noise* atau *outlier* dalam data yang dihasilkan oleh sensor cuaca perlu dihilangkan melalui teknik *data cleaning* yang kompleks. Selain itu, variasi dalam skala spasial dan temporal data, misalnya data radar lokal dibandingkan dengan data satelit global, memerlukan teknik interpolasi yang cermat untuk mencegah distorsi informasi (Raj et al., 2024).

Selain tantangan teknis, pengembangan model prediktif juga memerlukan algoritma yang mampu menangkap pola data kompleks dengan presisi tinggi. Model canggih, yang menggunakan *deep learning*, menunjukkan potensi besar dalam pemrosesan data cuaca temporal, tetapi membutuhkan data historis yang luas dan pelatihan intensif untuk memberikan hasil yang akurat (M et al., 2024).

Secara keseluruhan, meskipun data cuaca sangat penting untuk mendukung analisis dan prediksi kondisi atmosfer, berbagai tantangan dalam pengolahan data memerlukan pendekatan inovatif. Integrasi teknologi canggih seperti kecerdasan buatan, pembelajaran mesin, dan sistem komputasi berbasis awan diharapkan dapat mengatasi hambatan ini dan meningkatkan efisiensi serta akurasi prediksi cuaca di masa depan (Yao et al., 2022).

1.5.2 Machine Learning

Machine learning (ML) adalah cabang dari kecerdasan buatan (AI) yang fokus pada pengembangan algoritma yang memungkinkan komputer untuk belajar dan membuat prediksi berdasarkan data. Dalam konteks prediksi cuaca, ML digunakan untuk mengolah data cuaca yang kompleks, seperti suhu, kelembapan, curah hujan, dan pola angin. Berbagai algoritma ML, termasuk Linear Regression (LinReg), LightGBM Regressor (LGBM), K-Nearest Neighbors Regressor (KNN), Random Forest Regressor (RF), dan Support Vector Machines (SVM), sering digunakan untuk menangkap pola dalam data historis dan memberikan estimasi hasil berdasarkan data baru. Algoritma ini terbukti efektif dalam memanfaatkan korelasi non-linier dalam data atmosfer untuk meningkatkan akurasi prediksi (Maheswari & Gomathi, 2024; Raghuwanshi et al., 2024).

Salah satu kekuatan ML adalah kemampuannya untuk menangani volume data besar dan mempelajari pola dari parameter atmosfer yang saling terkait. Sebagai contoh, algoritma seperti Random Forest dan LightGBM telah menunjukkan keunggulan dalam menangkap pola kompleks dalam data cuaca dan memberikan hasil yang akurat bahkan dalam skenario data yang tidak lengkap atau bervariasi (Bartwal et al., 2024). Selain itu, integrasi algoritma seperti KNN dan SVM memberikan fleksibilitas tambahan untuk menyelesaikan masalah klasifikasi cuaca ekstrem, seperti mendeteksi potensi badai atau kekeringan berdasarkan pola data historis (Quinn & Abdelfattah, 2019).

Penggunaan ML dalam prediksi cuaca memberikan keuntungan tambahan berupa efisiensi komputasi yang lebih tinggi dibandingkan dengan pendekatan berbasis fisika tradisional, menjadikannya solusi yang menjanjikan untuk aplikasi modern seperti mitigasi bencana dan perencanaan sumber daya (Scher & Messori, 2018). Dengan kemajuan teknologi ini, prediksi cuaca tidak hanya menjadi lebih akurat tetapi juga lebih mudah diakses untuk aplikasi praktis.

Namun, meskipun ML mampu memodelkan hubungan linier dan non-linier dalam data cuaca, algoritma ini memiliki keterbatasan dalam menangkap pola temporal yang lebih kompleks dan ketergantungan antar variabel dalam jangka waktu panjang. Untuk mengatasi tantangan ini, *deep learning* (DL) hadir sebagai pendekatan yang lebih canggih, menawarkan kemampuan analisis data multivariat secara mendalam melalui jaringan saraf tiruan atau *neural networks* (Yang & Lu, 2022).

1.5.3 Deep Learning dan Temporal Fusion Transformer (TFT)

Deep learning (DL), sebagai cabang dari *machine learning* (ML), mengandalkan jaringan saraf tiruan (*neural networks*) yang terdiri dari banyak lapisan untuk mempelajari pola data yang kompleks. Dalam prediksi cuaca, DL unggul dalam menganalisis data temporal dan spasial. Model-model populer seperti Long Short-Term Memory (LSTM) dan Convolutional Neural Networks (CNN) sering digunakan untuk menangkap pola non-linear dan hubungan temporal dalam data atmosfer, seperti curah hujan ekstrem atau fluktuasi suhu yang mendadak (Yao et al., 2022).

Model LSTM dirancang untuk mengatasi masalah *vanishing gradients*, sehingga sangat efektif dalam menganalisis hubungan jangka panjang dalam data cuaca historis. Di sisi lain, CNN sangat baik untuk mengidentifikasi pola spasial, seperti distribusi awan atau curah hujan di wilayah tertentu, yang dapat diproses dengan data citra satelit atau radar (Shakya et al., 2020). Kedua model ini sering dikombinasikan untuk meningkatkan akurasi prediksi dengan memanfaatkan fitur temporal dan spasial secara simultan (Ren et al., 2021).

Temporal Fusion Transformer (TFT) kemudian dipilih karena kemampuannya, yang telah terbukti sebagai *State-of-the-Art*, dalam memprediksi data deret waktu multivariat dengan kompleksitas tinggi. Keunggulan TFT dibandingkan model lain, seperti LSTM dan CNN, terletak pada interpretabilitasnya yang mendalam, pengguna dapat memahami pengaruh setiap variabel terhadap prediksi secara langsung. Selain itu, kemampuannya untuk menangani data dengan elemen statis dan dinamis secara simultan menjadikannya sangat unggul, terutama dalam skenario yang membutuhkan prediksi akurat pada data yang kompleks, seperti data cuaca. Berikut adalah persamaan Temporal Fusion Transformer:

$$\hat{y}_i(q, t, \tau) = f_q(\tau, y_{i,t-k:t}, z_{i,t-k:t}, x_{i,t-k:t+\tau}, S_i) \quad (1)$$

Dengan keterangan sebagai berikut:

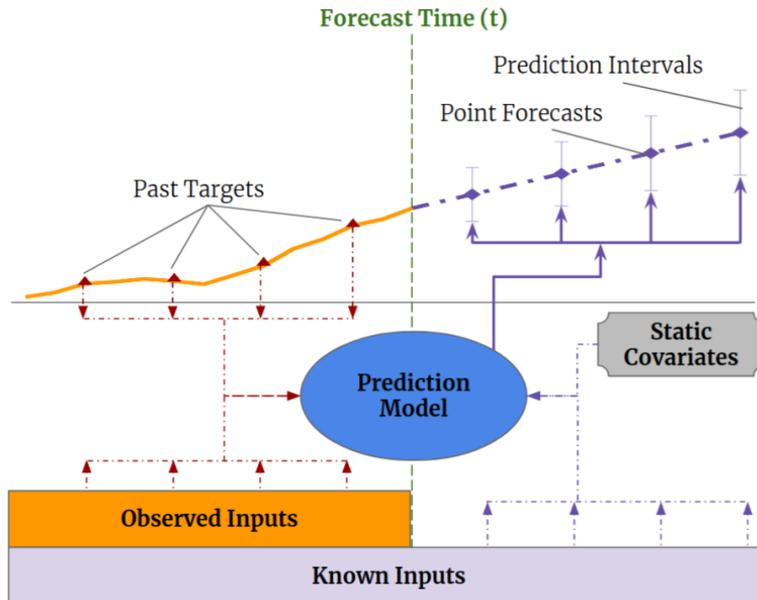
- q (*quantile*): persentil dalam distribusi probabilitas prediksi.
- t : waktu prediksi.
- τ : *horizon* atau waktu prediksi relatif (berapa jauh pada masa depan yang ingin diprediksi).

- k , panjang periode historis yang digunakan sebagai input.
- f_q : fungsi utama yang dipakai model untuk melakukan prediksi.
- $\hat{y}_{i,t-k:t}$: data historis dari target yang ingin diprediksi.
- $Z_{i,t-k:t}$: informasi input yang tidak diketahui sebelumnya.
- $X_{i,t:t+\tau}$: informasi input yang diketahui sebelumnya.
- S_i : informasi statis.

Berdasarkan publikasi resminya, Lim et al. menuliskan bahwa TFT memiliki kelebihan utama sebagai berikut:

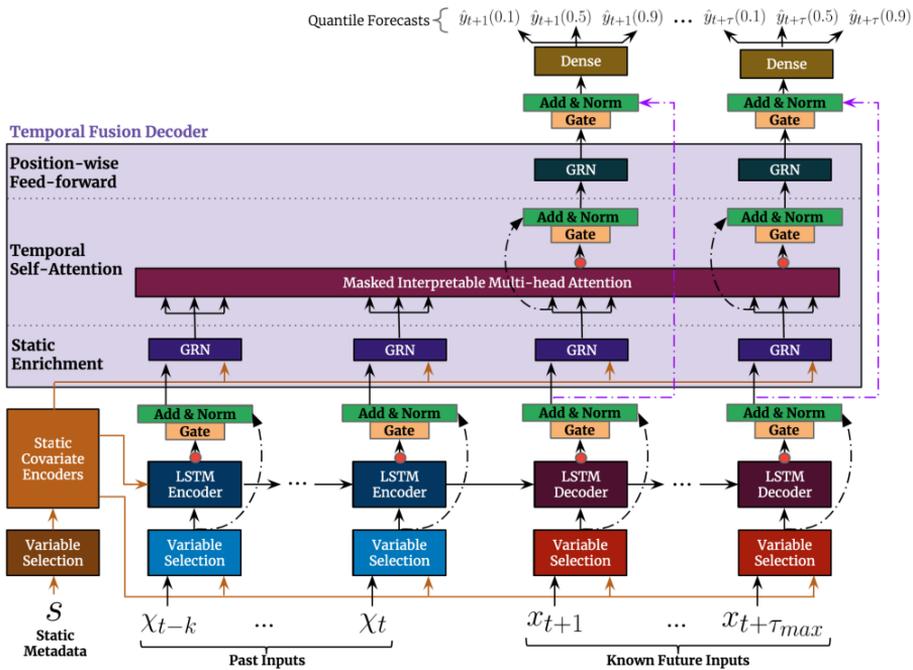
- ***Gating mechanisms***
Mekanisme gating untuk melewati komponen arsitektur yang tidak digunakan, sehingga memberikan kedalaman adaptif dan kompleksitas jaringan yang sesuai dengan berbagai jenis dataset dan skenario.
- ***Variable selection networks***
Jaringan pemilihan variabel untuk memilih variabel input yang relevan di setiap langkah waktu.
- ***Static covariate encoders***
Encoder kovariat statis untuk mengintegrasikan fitur statis ke dalam jaringan melalui pengkodean vektor konteks yang mengondisikan dinamika temporal.
- ***Temporal processing***
Pemrosesan temporal untuk mempelajari hubungan jangka panjang dan pendek dari input yang diamati dan yang sudah diketahui yang berubah seiring waktu. *Layer sequence-to-sequence* digunakan untuk pemrosesan lokal, sementara ketergantungan jangka panjang ditangkap menggunakan mekanisme *multi-head attention* yang baru dan dapat diinterpretasikan.
- ***Prediction intervals via quantile forecasts***
Interval prediksi melalui perkiraan kuantil untuk menentukan rentang nilai target yang mungkin di setiap horizon prediksi.

TFT menggabungkan fleksibilitas *attention mechanism* dengan kemampuan pembelajaran hubungan temporal dalam data. Hal ini menjadikannya sangat cocok untuk prediksi cuaca, berbagai variabel seperti suhu, tekanan, dan kecepatan angin harus dipertimbangkan secara simultan (Lim et al., 2021). Visualisasi input dari model TFT dapat dilihat pada **Gambar 1** di bawah.

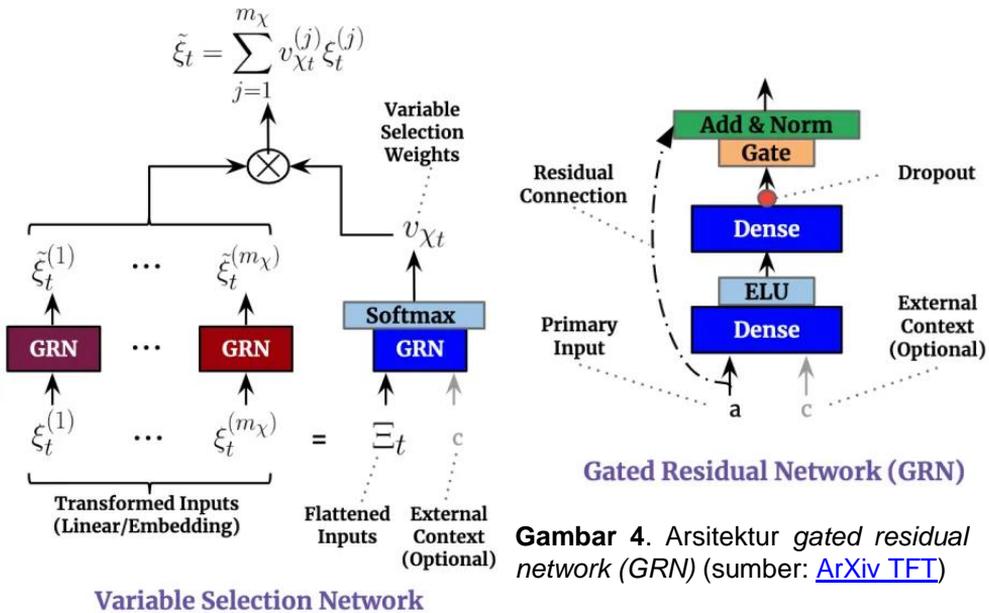


Gambar 1. Ilustrasi input model TFT (sumber: [ArXiv TFT](#))

Model TFT menggabungkan beberapa komponen utama seperti *encoder-decoder* untuk menangani data sekuensial, mekanisme *attention* untuk memprioritaskan variabel penting, serta komponen *gating* untuk mengontrol aliran informasi antara input dan *output*. Selain itu, model ini dilengkapi dengan kemampuan memproses data heterogen, baik data kategorikal maupun numerik, sehingga sangat cocok untuk analisis data cuaca. Untuk memahami lebih dalam mengenai model Temporal Fusion Transformer (TFT), arsitektur model TFT dapat dilihat pada **Gambar 2**, **Gambar 3**, dan **Gambar 4** di bawah.



Gambar 2. Arsitektur model Temporal Fusion Transformer (sumber: [ArXiv TFT](#))



Gambar 3. Arsitektur variable selection network (sumber: [ArXiv TFT](#))

Gambar 4. Arsitektur gated residual network (GRN) (sumber: [ArXiv TFT](#))

Arsitektur ini juga dirancang untuk memberikan interpretabilitas yang tinggi, memungkinkan analisis lebih lanjut terhadap faktor-faktor yang memengaruhi hasil prediksi.

TFT menawarkan peningkatan akurasi yang signifikan dibandingkan model tradisional dan bahkan beberapa model *deep learning* lainnya, seperti LSTM, CNN, atau GRU. Studi menunjukkan bahwa TFT tidak hanya memberikan prediksi yang lebih akurat tetapi juga lebih mudah diinterpretasikan oleh pengguna akhir, menjadikannya solusi yang sangat baik untuk analisis data cuaca kompleks (Burrichter et al., 2024). Selain itu, arsitektur TFT memungkinkan integrasi dengan berbagai jenis data masukan, seperti data satelit, radar, dan pengukuran lapangan, untuk memberikan wawasan yang lebih holistik tentang kondisi cuaca.

Integrasi *machine learning* dan *deep learning*, dalam kasus ini Temporal Fusion Transformer, dalam prediksi cuaca membuka peluang baru untuk meningkatkan akurasi dan interpretabilitas model. Dengan kemampuan untuk memproses data temporal multivariat dan memberikan wawasan tentang hubungan antar variabel, TFT memberikan solusi yang kuat untuk tantangan kompleks dalam analisis cuaca modern.

Namun, keunggulan prediksi yang ditawarkan oleh TFT tidak hanya bergantung pada akurasi dan interpretabilitas model, tetapi juga pada kemampuan untuk menyajikan hasil prediksi secara efektif kepada pengguna. Dalam konteks analisis data cuaca, hasil prediksi yang kompleks perlu dikomunikasikan dengan cara yang intuitif dan mudah dipahami, terutama oleh pengguna non-teknis seperti perencana kota, operator transportasi, atau pengambil keputusan di bidang mitigasi bencana. Di sinilah visualisasi data dan pengembangan dasbor interaktif memainkan peran penting, membantu menjembatani hasil prediksi dari model seperti TFT ke dalam wawasan yang dapat ditindaklanjuti.

1.5.4 Visualisasi Data dan Dasbor Interaktif

Visualisasi data adalah proses mengubah data mentah menjadi representasi visual yang informatif, seperti grafik, peta, atau diagram, untuk mempermudah pemahaman pola, tren, dan hubungan antar data. Dalam analisis data cuaca, visualisasi sangat penting untuk mengkomunikasikan informasi kompleks, seperti fluktuasi suhu, curah hujan, dan arah angin, kepada pengguna akhir secara intuitif (Suresh & P, 2022).

Visualisasi membantu menjembatani kompleksitas hasil model prediksi dengan kebutuhan pengguna untuk mengambil keputusan berbasis data. Misalnya:

- **Grafik temporal**, memvisualisasikan tren suhu, curah hujan, atau angin secara linier dari waktu ke waktu untuk mendeteksi pola musiman.

- **Heatmap**, menyajikan distribusi curah hujan atau suhu pada peta geografis untuk memahami pola spasial.
- **Grafik interaktif**, memberikan pengguna kemampuan untuk mengeksplorasi data prediksi dengan menyoroti detail pada waktu atau lokasi tertentu.

Visualisasi interaktif ini memungkinkan pengguna untuk lebih memahami hasil prediksi yang dihasilkan oleh model seperti Temporal Fusion Transformer (TFT), sehingga meningkatkan kepercayaan dan efektivitas penggunaan hasil analisis (Cohen et al., 2021).

Dasbor interaktif adalah platform yang dirancang untuk menyajikan data secara dinamis dan mudah diakses oleh pengguna. Dalam konteks data cuaca, dasbor ini memungkinkan integrasi berbagai elemen, seperti:

- **Data historis**, menampilkan data cuaca masa lalu untuk membantu memahami pola.
- **Prediksi cuaca**, menyajikan hasil prediksi model TFT dengan visualisasi yang menarik dan informatif.
- **Analisis interaktif**, pengguna dapat menyesuaikan tampilan berdasarkan parameter tertentu, seperti wilayah, periode waktu, atau jenis data cuaca.

Platform ini dapat dibangun menggunakan teknologi seperti Streamlit, Dash, atau Tableau, yang mendukung integrasi data *real-time* dengan antarmuka yang ramah pengguna (Hao et al., 2007).

Manfaat visualisasi data dan dasbor interaktif dalam prediksi cuaca sangat beragam. **Pertama**, interpretabilitas hasil prediksi menjadi lebih mudah dipahami, bahkan untuk model prediksi yang kompleks seperti TFT. Ini berarti bahwa wawasan yang dihasilkan dari model-model tersebut dapat diterjemahkan menjadi informasi yang lebih mudah diakses dan dimengerti oleh berbagai pengguna. **Kedua**, penggunaan dasbor memfasilitasi pengambilan keputusan yang cepat dengan memberikan akses *real-time* ke informasi penting. Hal ini sangat krusial dalam situasi seperti mitigasi bencana atau perencanaan logistik, keputusan cepat dan akurat sangat dibutuhkan. **Terakhir**, visualisasi interaktif meningkatkan keterlibatan pengguna, terutama bagi mereka yang non-teknis. Mereka dapat mengeksplorasi data dengan cara yang lebih intuitif, sehingga pemahaman dan keterlibatan mereka terhadap informasi yang disajikan meningkat (Leffrang & Muller, 2021).

Dalam penelitian ini, dasbor interaktif dibangun untuk memvisualisasikan hasil prediksi cuaca Kota Makassar yang dihasilkan oleh model Temporal Fusion Transformer. Dengan dasbor ini, pengguna dapat memantau prediksi cuaca, menganalisis pola historis, dan mengeksplorasi tren lokal secara interaktif. Sistem ini dirancang untuk menjadi alat yang mudah diakses, baik oleh pengguna teknis maupun non-teknis, untuk mendukung pengambilan keputusan berbasis data.

Dalam memvisualisasikan data yang informatif dan membangun dasbor interaktif yang efektif diperlukan pemilihan *framework* dan teknologi yang sesuai. *Framework* yang digunakan harus mampu mendukung pemrosesan data besar, integrasi model prediktif seperti Temporal Fusion Transformer (TFT), serta menghasilkan antarmuka visual yang dinamis dan interaktif. Oleh karena itu, dalam penelitian ini, dipilih berbagai teknologi yang tidak hanya mendukung implementasi model dan visualisasi, tetapi juga memastikan sistem dapat berfungsi secara efisien dan mudah diakses oleh pengguna.

1.5.5 *Framework* dan Teknologi yang Digunakan

Dalam penelitian ini, berbagai *framework* dan teknologi digunakan untuk memastikan efisiensi dalam pemrosesan data, pengembangan model, dan visualisasi prediksi. Pemilihan teknologi ini mempertimbangkan kompatibilitas dengan Temporal Fusion Transformer (TFT) serta kemudahan integrasinya dalam membangun dasbor interaktif untuk prediksi cuaca di Kota Makassar.

Framework untuk pemrosesan data:

- **NumPy**
NumPy adalah library fundamental dalam Python untuk komputasi numerik. Library ini digunakan untuk manipulasi array multidimensi dan operasi matematis, seperti normalisasi data dan perhitungan statistik, yang diperlukan dalam *preprocessing* data cuaca.
- **Pandas**
Pandas memungkinkan pengelolaan data berbentuk tabular dengan fungsi seperti manipulasi, penggabungan, dan pengisian data yang hilang. Dalam konteks ini, Pandas digunakan untuk mengolah *dataset* cuaca, seperti memproses data historis suhu dan curah hujan.

Framework untuk visualisasi data dan dasbor:

- **Streamlit**
Streamlit adalah *framework open-source* yang digunakan untuk membangun aplikasi web berbasis Python secara cepat dan interaktif. Dalam penelitian ini, Streamlit digunakan untuk mengembangkan dasbor yang menampilkan hasil prediksi cuaca secara *real-time*, memungkinkan pengguna untuk mengeksplorasi data secara dinamis.
- **Matplotlib dan PyPlot**
Matplotlib, khususnya modul PyPlot, digunakan untuk membuat grafik statis seperti *line chart* dan *scatter plot*. Grafik ini mempermudah visualisasi pola data cuaca, seperti tren curah hujan atau suhu harian.

Framework untuk pengembangan model:

- **PyTorch dan PyTorch Lightning**
PyTorch adalah library *deep learning* yang fleksibel, digunakan untuk membangun dan melatih model Temporal Fusion Transformer (TFT). PyTorch Lightning menyederhanakan pengelolaan kode dan eksperimen pelatihan model, memungkinkan pengembangan yang lebih terstruktur dan efisien.
- **Nixtla**
Nixtla adalah library khusus untuk *forecasting* yang memfasilitasi integrasi berbagai model deret waktu. Library ini memberikan dukungan untuk *preprocessing* data dan *pipeline* yang efisien dalam eksperimen *forecasting*.

Dengan kombinasi *framework* ini, penelitian ini mencakup semua aspek mulai dari pengumpulan data cuaca, pelatihan model prediksi menggunakan PyTorch dan PyTorch Lightning yang dibungkus dalam Nixtla, hingga visualisasi hasil di dasbor interaktif berbasis Streamlit. Penelitian ini juga menggunakan basis data MySQL untuk menyimpan data cuaca dan hasil prediksi model. Penggunaan Nixtla juga memastikan model TFT dapat dikembangkan, divalidasi, dan dianalisis secara menyeluruh.

BAB II METODE

Pada bab ini, dijelaskan secara rinci prosedur, analisis, dan perhitungan yang digunakan dalam penelitian ini. Penelitian ini berfokus pada pengembangan dasbor interaktif untuk analisis dan prediksi data cuaca di Kota Makassar menggunakan algoritma Temporal Fusion Transformer (TFT). Metode penelitian mencakup pengumpulan data, *preprocessing* data, implementasi algoritma, pembangunan dasbor, serta evaluasi kinerja sistem.

2.1 Waktu dan Lokasi Penelitian

Penelitian ini dilaksanakan di Makassar selama kurang lebih tiga bulan, dari bulan November 2024 sampai bulan Januari 2025 seperti yang dapat dilihat pada **Tabel 1**.

Tabel 1. Jadwal kegiatan penelitian.

No.	Tahap	November 2024				Desember 2024				Januari 2025			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan Data Cuaca	■	■	■	■								
2	Pelatihan Model				■	■	■	■					
3	Pembuatan Dasbor					■	■	■	■	■			
4	Integrasi Data dan Model ke Dasbor							■	■	■	■	■	
5	Evaluasi Model dan Sistem										■	■	■
6	Deployment												■

2.2 Pengumpulan Data

Penelitian ini menggunakan data cuaca Kota Makassar yang diperoleh dari Open-Meteo. Open-Meteo merupakan layanan API cuaca terbuka yang menyediakan data meteorologi global dari berbagai sumber resmi dan model tepercaya (Zippenfenig, 2024). Data diperoleh melalui pemanggilan API dengan parameter sebagai berikut:

- **Lokasi Makassar**, dengan garis lintang -5.1486 dan garis bujur 119.4319.
- **Zona waktu Singapura**, GMT+8.
- **Variabel-variabel cuaca harian**, seperti kode cuaca, temperatur maksimal, temperatur minimal, dll.

Data awal mencakup periode 1 Januari 2000 – 31 Desember 2024, dengan jumlah data sebanyak 9132 entri. Variabel-variabel yang ada pada data beserta penjelasannya dijelaskan pada **Tabel 2** di bawah.

Tabel 2. Variabel data cuaca.

Variabel	Unit	Deskripsi
time	HH-BB- TTTT	Tanggal
temperature_2m_max temperature_2m_min temperature_2m_mean	°C (°F)	Suhu udara maksimum, minimum, dan rata-rata harian pada ketinggian 2 meter di atas tanah.
apparent_temperature_max apparent_temperature_min apparent_temperature_mean	°C (°F)	Suhu terasa maksimum, minimum, dan rata-rata harian.
precipitation_sum	mm	Jumlah total presipitasi harian (termasuk hujan, gerimis, dan salju).
rain_sum	mm	Total hujan harian.
precipitation_hours	Jam	Jumlah jam dengan hujan.
sunshine_duration	Detik	Jumlah detik sinar matahari per hari, dihitung berdasarkan iradiasi langsung yang melebihi 120 W/m ² , sesuai definisi WMO. Durasi sinar matahari akan selalu lebih pendek dibandingkan durasi siang hari karena adanya waktu fajar dan senja.
daylight_duration	Detik	Jumlah detik siang hari per hari.
wind_speed_10m_max wind_gusts_10m_max	km/j (mph, m/s, knots)	Kecepatan angin maksimum dan hembusan angin pada suatu hari.
shortwave_radiation_sum	MJ/m ²	Jumlah radiasi matahari pada suatu hari dalam Megajoule.
et0_fao_evapotranspiration	mm	Total harian ET ₀ Evapotranspirasi Referensi pada ladang rumput yang terairi dengan baik.

Pada penelitian ini, ***precipitation_sum*** dijadikan sebagai variabel target dan variabel numerik lainnya dijadikan sebagai fitur historis (*past covariate*). Data disimpan dalam basis data MySQL untuk kebutuhan pengolahan lebih lanjut. Struktur dari basis data yang digunakan dapat dilihat pada **Gambar 5**.

Weather Data	Predictions
time	id
temperature_2m_max	model
temperature_2m_min	hasil_forecast
temperature_2m_mean	insample
apparent_temperature_max	interpretation
apparent_temperature_min	metrics
apparent_temperature_mean	timestamp
daylight_duration	
sunshine_duration	
precipitation_sum	
rain_sum	
precipitation_hours	
wind_speed_10m_max	
wind_gusts_10m_max	
shortwave_radiation_sum	
et0_fao_evapotranspiration	

Gambar 5. Struktur tabel pada basis data.

2.3 Implementasi Algoritma Temporal Fusion Transformer (TFT)

Penelitian ini menggunakan library Nixtla, khususnya *neuralforecast*, untuk mengimplementasikan algoritma Temporal Fusion Transformer (TFT). Nixtla menyediakan abstraksi tingkat tinggi yang memudahkan penggunaan berbagai model *time series*, termasuk TFT, dengan *preprocessing* otomatis. Proses *preprocessing* ini diotomatiskan oleh Nixtla, yang memungkinkan peneliti untuk lebih fokus pada konfigurasi model dan analisis hasil, daripada *preprocessing* data secara manual.

Dengan menggunakan library Nixtla, implementasi model TFT menjadi lebih efisien dan mudah. Nixtla menangani banyak detail implementasi dan memungkinkan peneliti untuk lebih fokus pada konfigurasi model dan analisis hasil.

2.3.1 Konfigurasi Model TFT dengan Nixtla

Setelah data di-*preprocess* secara otomatis oleh Nixtla, model TFT dikonfigurasi dengan parameter-parameter yang relevan. Konfigurasi ini dilakukan melalui fungsi TFT yang disediakan oleh Nixtla. Parameter-parameter utama yang dikonfigurasi pada model TFT dapat dilihat pada **Tabel 3** di bawah.

Tabel 3. *Hyperparameter* model.

Hyperparameter	Nilai
input_size	60
horizon	30
hidden_size	128
n_head	4
dropout	0.1
learning_rate	0.001
loss_fn	HuberMQLoss
batch_size	32
epoch	1000
early_stop_patience_steps	10
grn_activation	ELU
val_check_steps	100
valid_loss	MQLoss

2.3.2 Pelatihan dan Validasi Model

Model TFT dilatih menggunakan fungsi *cross_validation*, yaitu fungsi yang disediakan oleh Nixtla untuk melakukan validasi silang, sehingga memungkinkan evaluasi model yang lebih *robust*. Data yang digunakan dalam pelatihan merupakan data historis yang telah melalui proses *preprocessing* secara otomatis. Proses pelatihan menggunakan HuberMQLoss untuk menghitung *training loss* dan MQLoss untuk menghitung *validation loss*, sehingga model dapat dievaluasi dengan lebih akurat.

Validasi silang dilakukan dengan membagi dataset menjadi tiga bagian utama, yaitu data latih, data validasi, dan data uji. Data validasi terdiri dari 90 hari terakhir sebelum periode pengujian, sedangkan data uji terdiri dari 90 hari terakhir dalam dataset. Perhitungan metrik evaluasi pada data validasi dan data uji dilakukan menggunakan teknik *sliding window*, mengingat model hanya mampu memprediksi 30 hari ke depan dalam satu iterasi (*horizon* = 30). Dengan pendekatan ini, model diuji pada data yang belum pernah dilihat sebelumnya, sehingga dapat memberikan gambaran yang lebih akurat terkait performanya dalam melakukan prediksi.

Selain itu, pelatihan model menggunakan strategi penghentian dini (*early stopping*) untuk mencegah *overfitting*. Proses pelatihan akan dihentikan secara otomatis jika performa model pada data validasi tidak menunjukkan peningkatan selama 10 *epoch*. Hal ini bertujuan untuk mengoptimalkan kinerja model tanpa melakukan pelatihan yang berlebihan, sehingga model tetap dapat melakukan generalisasi dengan baik pada data baru.

2.3.3 Interpretasi Model dengan Nixtla

Salah satu keunggulan Nixtla adalah kemampuannya untuk menyediakan interpretasi model. Melalui fungsi bawaan TFT yang disediakan Nixtla, kita dapat mengekstrak beberapa informasi interpretasi, di antaranya:

- **Attention weights**, untuk memvisualisasikan bagaimana model TFT fokus pada data yang disediakan, baik itu statis, lampau, dan data yang tersedia pada masa depan.
- **Feature importance**, untuk menunjukkan variabel-variabel yang paling berpengaruh dalam prediksi.
- **Feature importance correlation**, untuk menunjukkan seberapa besar pengaruh antar variabel dalam prediksi.

Interpretasi ini sangat penting untuk memahami bagaimana model bekerja dan memberikan *insight* yang berharga.

2.3.4 Prediksi dengan Nixtla

Setelah model dilatih, Nixtla menyediakan fungsi *predict* untuk menghasilkan prediksi pada data yang baru. Hasil prediksi mencakup nilai *median* dan *quantile* lainnya.

2.4 Pembuatan Dasbor Interaktif

Dasbor interaktif dalam penelitian ini dirancang menggunakan Streamlit, sebuah *framework* Python yang *open-source* untuk membuat aplikasi web berbasis data. Streamlit dipilih karena kemudahan penggunaannya, sintaks yang intuitif, dan kemampuannya untuk membuat aplikasi web yang interaktif dan responsif dengan cepat tanpa memerlukan keahlian *web development* yang mendalam. Proses pembangunan dasbor meliputi beberapa tahapan dan fitur utama:

2.4.1 Desain Antarmuka Pengguna (UI)

Antarmuka pengguna (UI) dasbor dirancang agar intuitif, mudah digunakan, dan informatif. Beberapa elemen desain UI utama meliputi:

Halaman utama. Menampilkan informasi cuaca terkini berdasarkan kelurahan yang dipilih. Tampilan utama terdiri dari metrik cuaca seperti suhu, kelembapan, kecepatan angin, dan curah hujan dengan tampilan yang *real-time*, serta informasi perubahan nilai dalam periode tertentu. Selain itu, pada halaman utama juga terdapat tombol navigasi untuk menuju halaman lain seperti Dataset, Analisis, dan Forecast.

Halaman dataset. Menampilkan data historis cuaca. Halaman ini menampilkan informasi seperti jumlah data, periode data, jumlah kolom, statistik deskriptif, informasi kolom, serta menampilkan data mentah dalam bentuk tabel dengan opsi *download* data dalam format CSV.

Halaman analisis. Menampilkan analisis data cuaca. Halaman ini bertujuan untuk membantu pengguna menganalisis data cuaca historis. Halaman ini menampilkan informasi analisis musiman, tren tahunan, kejadian ekstrem, analisis detail (distribusi data dan *time series* untuk setiap variabel), matriks korelasi, dan identifikasi anomali.

Halaman forecast. Menampilkan prakiraan cuaca dari BMKG dan hasil prediksi dari model *deep learning* (DL). Untuk halaman prakiraan BMKG, informasi yang ditampilkan adalah kondisi cuaca saat ini beserta metrik terkaitnya, tren cuaca (suhu, kelembapan, dan curah hujan), serta data prakiraan secara lengkap. Sedangkan untuk halaman prakiraan DL yaitu grafik hasil prediksi model TFT, metrik evaluasi, visualisasi dan analisis dari interpretasi model TFT, seperti *attention weights*, *feature importance*, dan korelasi variabel.

Halaman bantuan. Menampilkan informasi dasar dan panduan penggunaan aplikasi.

2.4.2 Integrasi Data

Data historis dan hasil prediksi dari model TFT diintegrasikan ke dalam dasbor secara *real-time*. Proses integrasi data dilakukan dengan cara:

Pengambilan data historis. Data historis cuaca diambil dari basis data (MySQL) menggunakan koneksi yang telah dikonfigurasi pada aplikasi Streamlit. Data ini kemudian di-*preprocess* dan ditampilkan dalam bentuk tabel atau grafik.

Pengambilan data prakiraan BMKG. Data prakiraan cuaca BMKG diambil langsung melalui API dan data disajikan secara *realtime* pada *dashboard*.

Integrasi data prediksi TFT. Hasil prediksi dari model TFT yang telah dilatih (menggunakan *library* *neuralforecast*) diintegrasikan langsung ke dalam dasbor. Data prediksi ini diambil dari basis data (MySQL). Model melakukan *forecast* secara berkala dan hasil prediksinya disimpan pada basis data.

Visualisasi data. Data historis dan data prediksi ditampilkan dalam berbagai bentuk visualisasi yang telah disediakan Plotly seperti grafik *line*, grafik *bar*, dan *heatmap*.

Pembaruan data. Data diperbarui secara berkala sesuai dengan kebutuhan. Data BMKG diperbarui secara berkala dengan interval beberapa jam sekali dan data prediksi dengan interval harian. Data historis diperbarui setiap dua hari sekali (sesuai dokumentasi Open-Meteo), dan data dari API di-*fetch* setiap hari menggunakan *cron jobs* untuk mengecek apakah ada data baru atau tidak kemudian disimpan ke dalam basis data MySQL. Streamlit juga melakukan *re-render* halaman secara otomatis setiap kali ada perubahan pada *state* aplikasi.

2.4.3 Fitur Dasbor

Berdasarkan penjelasan di atas, pada penelitian ini, dasbor akan menyediakan fitur-fitur berikut:

- **Pemfilteran data**, pengguna dapat memfilter data berdasarkan rentang waktu, lokasi (kecamatan dan kelurahan), dan variabel cuaca.
- **Visualisasi data**, menampilkan visualisasi data historis, tren, dan prediksi dalam bentuk grafik.
- **Prakiraan cuaca**, menampilkan prakiraan cuaca dari BMKG dan hasil prediksi model TFT.
- **Interpretasi model**, menampilkan hasil analisis dari interpretasi model TFT.
- **Navigasi**, dasbor mudah dinavigasi dengan adanya menu di *sidebar* dan *tab* pada halaman forecast.
- **Akses data**, pengguna dapat mengunduh data dalam format CSV.

2.5 Evaluasi Sistem

Evaluasi sistem dalam penelitian ini dilakukan untuk mengukur kinerja model prediksi dan fungsionalitas dasbor interaktif secara keseluruhan. Evaluasi ini mencakup dua aspek utama, yaitu evaluasi model prediksi dan evaluasi dasbor interaktif.

2.5.1 Evaluasi Model Prediksi

Kinerja model prediksi Temporal Fusion Transformer (TFT) dievaluasi menggunakan metrik-metrik berikut:

- **Mean Absolute Error (MAE)**

Mengukur rata-rata selisih absolut antara nilai aktual (y) dan nilai prediksi (\hat{y}). MAE memberikan gambaran seberapa besar rata-rata error yang dihasilkan model dalam satuan yang sama dengan data target. Simbol n menunjukkan jumlah seluruh data. Berikut rumus dari MAE:

$$MAE = \left(\frac{1}{n}\right) \sum |y_i - \hat{y}_i| \quad (2)$$

- **Mean Squared Error (MSE)**

Mengukur rata-rata kuadrat selisih antara nilai aktual (y) dan nilai prediksi (\hat{y}). MSE memberikan bobot lebih besar pada error yang besar dan nilainya berkisar antara 0 hingga tak terhingga. Simbol n menunjukkan jumlah seluruh data. Berikut rumus dari MSE:

$$MSE = \left(\frac{1}{n}\right) \sum (y_i - \hat{y}_i)^2 \quad (3)$$

- **Root Mean Squared Error (RMSE)**

Mengukur akar kuadrat dari rata-rata kuadrat selisih antara nilai aktual (y) dan nilai prediksi (\hat{y}). Simbol n menunjukkan jumlah seluruh data. Berikut rumus dari RMSE:

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum (y_i - \hat{y}_i)^2} \quad (4)$$

- **Mean Absolute Percentage Error (MAPE)**

Mengukur rata-rata persentase error absolut antara nilai aktual (y) dan nilai prediksi (\hat{y}). MAPE memberikan gambaran seberapa besar rata-rata error dalam bentuk persentase relatif terhadap nilai aktual. Simbol n menunjukkan jumlah seluruh data. Berikut rumus dari MAPE:

$$MAPE = \left(\frac{1}{n}\right) \sum \left| \frac{(y_i - \hat{y}_i)}{y_i} \right| \times 100\% \quad (5)$$

- **Symmetric Mean Absolute Percentage Error (SMAPE)**

Mengukur rata-rata persentase error absolut simetris antara nilai aktual (y) dan nilai prediksi (\hat{y}). SMAPE memberikan persentase error yang dinormalisasi, dan hasil rentang antara 0-100%. SMAPE juga lebih stabil ketika nilai aktualnya mendekati nol, tidak seperti MAPE. Simbol n menunjukkan jumlah seluruh data. Berikut rumus dari SMAPE:

$$SMAPE = \left(\frac{1}{n}\right) \sum \left(\frac{|\hat{y}_i - y_i|}{(|y_i| + |\hat{y}_i|)} \right) \times 2 \times 100\% \quad (6)$$

2.5.2 Evaluasi Dasbor Interaktif

Evaluasi dilakukan menggunakan pendekatan pengujian *blackbox*, yang berfokus pada pengujian fungsionalitas sistem dari sudut pandang pengguna akhir tanpa melihat kode internal. Pendekatan ini dipilih untuk memastikan bahwa dasbor dapat digunakan secara efektif oleh pengguna dari berbagai latar belakang teknis.

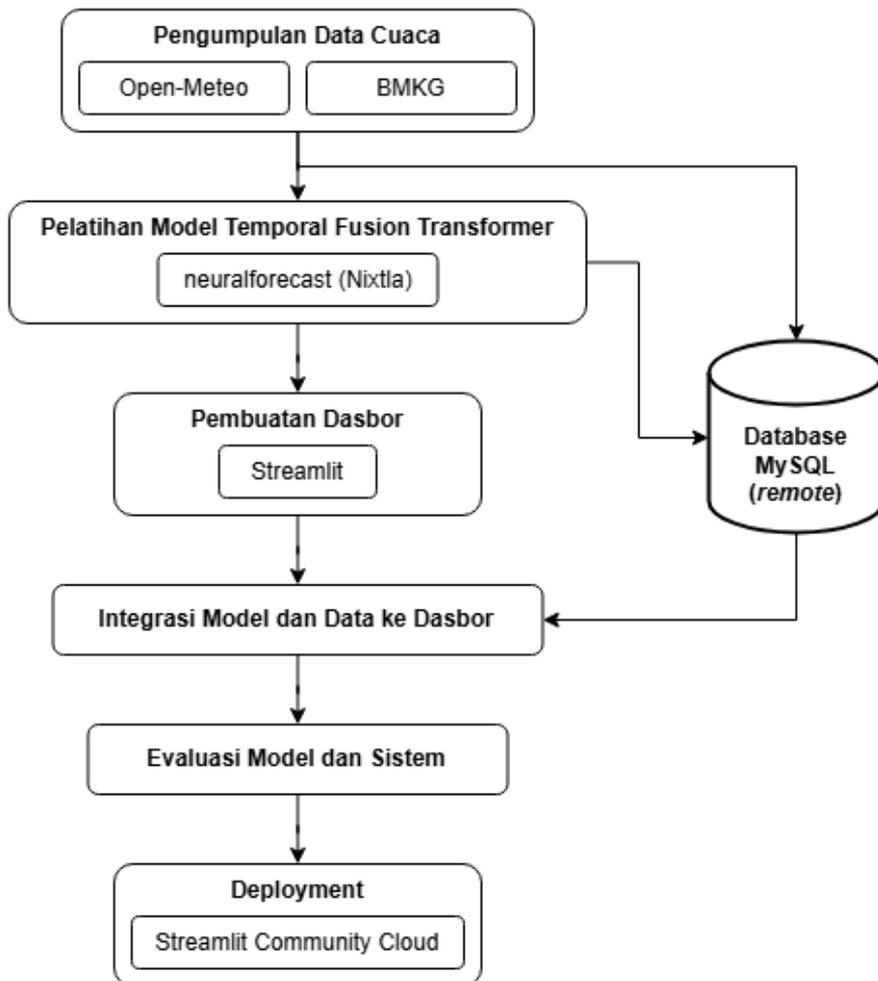
Pengujian dilakukan pada berbagai perangkat dan konfigurasi, termasuk pada ukuran layar Full-HD (resolusi 1920x1080), HD (resolusi 1366x768), tablet, dan *smartphone*, menggunakan *browser* Chrome versi terbaru, dengan koneksi internet stabil minimal 10 Mbps. Kasus uji dirancang untuk mencakup beberapa aspek utama dasbor, termasuk:

- Pengujian halaman utama untuk memastikan pemuatan halaman dan tampilan data *real-time* yang akurat.
- Pengujian halaman dataset untuk memastikan pemuatan, pemfilteran, dan pengunduhan data historis yang benar.
- Pengujian halaman analisis untuk memastikan pemuatan grafik musiman dan tren tahunan, serta anomali yang akurat.
- Pengujian halaman prakiraan BMKG dan *deep learning* untuk memastikan data prakiraan termuat dan visualisasi interpretasi model dapat diakses.
- Pengujian responsivitas untuk memastikan tata letak dasbor beradaptasi dengan baik pada berbagai ukuran layar.
- Pengujian interaktivitas untuk memastikan bahwa pengguna dapat berinteraksi dengan data secara fleksibel dan mendapatkan pengalaman yang optimal.

Keberhasilan pengujian akan diukur menggunakan keberhasilan dan waktu pemuatan halaman.

2.6 Alur Penelitian

Ringkasan alur pembuatan dasbor yang akan dilakukan dapat dilihat pada **Gambar 6** di bawah:



Gambar 6. Alur pembuatan dasbor.