

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri ritel membutuhkan prediksi penjualan akurat untuk mendukung pengambilan keputusan terkait persediaan, distribusi, dan promosi sesuai perubahan permintaan pasar. Kebutuhan tersebut sejalan dengan pertumbuhan ritel Indonesia pada 2023 yang berada pada kisaran 4% hingga 4,2%, serta proyeksi Indeks Penjualan Riil November 2025 yang meningkat 5,9% secara tahunan [1, 2]. Perubahan perilaku belanja masyarakat juga menunjukkan transformasi pasar, ditandai dengan kenaikan rata-rata proporsi pengguna internet yang berbelanja secara digital dari 53% pada 2019 menjadi 60% pada 2020 hingga 2021 [3]. Perkembangan tersebut menunjukkan bahwa pola permintaan ritel dipengaruhi oleh perubahan kanal transaksi dan intensitas konsumsi. Kesalahan prediksi dalam situasi tersebut dapat memicu kelebihan stok, kekurangan stok, dan inefisiensi operasional, sehingga model prediksi penjualan perlu mempertimbangkan faktor eksternal selain histori penjualan.

Kebutuhan untuk mempertimbangkan faktor eksternal tersebut menjadi dasar penggunaan dataset *Walmart Sales Forecast* dalam penelitian ini. Dataset *Walmart Sales Forecast* digunakan karena menyediakan data penjualan mingguan dari 45 toko yang dilengkapi variabel eksternal, meliputi *Temperature*, *IsHoliday*, dan *Markdown1-5*. Pada penelitian ini, faktor cuaca dibatasi pada variabel *Temperature* yang merepresentasikan suhu udara pada wilayah toko. Dengan Batasan tersebut, cuaca tidak dimaknai sebagai seluruh kondisi meteorologis seperti curah hujan, kelembapan, kecepatan angin, salju atau kondisi cuaca ekstrem, melainkan sebagai indikator suhu udara yang tersedia dalam dataset. Penelitian terdahulu menunjukkan bahwa hubungan cuaca terhadap penjualan ritel tidak selalu seragam, tetapi bergantung pada karakter wilayah, konsep penjualan, dan tingkat perubahan suhu yang terjadi [4, 5]. Variabel *IsHoliday* merepresentasikan hari besar dan hari libur yang dapat mengubah pola pembelian mingguan melalui peningkatan transaksi sebelum, selama, maupun setelah periode perayaan. Variabel *Markdown1-5* merepresentasikan intensitas promosi yang dapat memengaruhi nilai belanja, jumlah item yang dibeli, dan jarak waktu antartransaksi pelanggan [6, 7]. Keterkaitan ketiga variabel tersebut menunjukkan bahwa fluktuasi penjualan mingguan dapat dianalisis melalui hubungan antara suhu udara, kalender libur, dan promosi yang terjadi pada periode yang sama.

Karakteristik penjualan ritel mingguan yang berkaitan dengan suhu udara, hari besar dan hari libur, serta promosi menunjukkan bahwa pola permintaan tidak hanya bergerak mengikuti histori penjualan. Pola tersebut dapat terbentuk melalui kombinasi tren, musiman, perubahan perilaku belanja pada periode tertentu, hubungan nonlinier antarvariabel, serta ketergantungan nilai penjualan dari waktu ke waktu. Kondisi ini menunjukkan bahwa model tunggal perlu diuji terlebih dahulu untuk melihat kemampuan masing-masing pendekatan dalam menangkap pola penjualan. Prophet relevan digunakan untuk merepresentasikan tren, pola musiman, dan efek kalender secara eksplisit, sedangkan *LSTM* memiliki kemampuan dalam mempelajari dependensi sekuensial pada data deret waktu [8]. Pada pendekatan lain, *XGBoost* dinilai relevan untuk peramalan penjualan ritel karena mampu menghasilkan kinerja kompetitif pada data penjualan terstruktur [9]. Perbedaan karakteristik tersebut menunjukkan bahwa pendekatan *hybrid machine learning* diperlukan untuk menggabungkan kekuatan model yang saling melengkapi, khususnya dalam merepresentasikan pola kalender, dinamika urutan waktu, dan relasi non-linear antarvariabel.

Penggunaan pendekatan *hybrid* dalam penelitian terdahulu memberikan dasar empiris bagi pengembangan model prediksi penjualan ritel, meskipun hasilnya tetap perlu diuji kembali sesuai karakteristik dataset yang digunakan. Kombinasi *Prophet-LightGBM* pada penjualan komoditas ritel menghasilkan nilai R^2 sebesar 0,85, yang menunjukkan bahwa pemodelan tren, musiman, dan hari libur melalui *Prophet* dapat dilengkapi dengan kemampuan *LightGBM* dalam mempelajari fitur multidimensi [5]. Pada penelitian *e-commerce*, model *LSTM-Prophet* menurunkan *average absolute error* menjadi 8,7% lebih rendah dibandingkan *Prophet* tunggal sebesar 12,3% dan *LSTM* tunggal sebesar 10,5% [8]. Sementara itu, *XGBoost* telah diterapkan pada peramalan penjualan ritel *UB Mart* dengan rata-rata *MAE* sebesar 11,58 dan *RMSE* sebesar 29,19 [9]. Kombinasi *XGBoost-LSTM* juga menunjukkan bahwa integrasi model berbasis *boosting* dan model sekuensial dapat digunakan untuk meningkatkan kemampuan prediksi pada data deret waktu penjualan [10]. Temuan tersebut menunjukkan bahwa pola penjualan ritel perlu direpresentasikan melalui pendekatan yang mempertimbangkan tren, pola musiman, hubungan non-linear, dan dependensi waktu secara terpadu. Namun, karena setiap penelitian memiliki karakteristik data yang berbeda, penelitian ini tetap melakukan pengujian langsung terhadap model tunggal dan model *hybrid* pada dataset *Walmart Sales Forecast*.

Sejalan dengan kebutuhan perbandingan performa yang objektif, penelitian diarahkan pada evaluasi tiga pasangan model *hybrid*, yakni *XGBoost-LSTM*, *Prophet-XGBoost*, dan

Prophet-LSTM. Perbandingan tersebut dilakukan untuk menilai model yang paling sesuai bagi prediksi penjualan ritel mingguan dengan mempertimbangkan faktor cuaca yang direpresentasikan oleh suhu udara, hari libur dan hari besar, serta aktivitas promosi. Agar perbandingan performa model dilakukan secara objektif, optimasi *hyperparameter* diterapkan melalui *Grid Search* karena pendekatan tersebut menelusuri kombinasi parameter secara menyeluruh pada ruang kandidat yang telah ditetapkan [11]. Pilihan tersebut relevan ketika tujuan utama penelitian terletak pada perbandingan performa model secara adil, transparan, dan mudah direplikasi. Dengan demikian, setiap model dapat memperoleh konfigurasi parameter terbaik berdasarkan ruang pencarian yang sama-sama terdefinisi, sehingga hasil perbandingan performa model tunggal dan model hybrid menjadi lebih terukur. Kerangka CRISP-DM digunakan agar proses penelitian tersusun secara sistematis melalui tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [9, 12]. Melalui kerangka tersebut, identifikasi masalah, pemilihan variabel eksternal, penyusunan proses pemodelan, dan evaluasi hasil prediksi dapat dilakukan secara terarah serta memiliki dasar metodologis yang jelas. Berdasarkan uraian tersebut, penelitian ini berjudul **“Prediksi Penjualan Ritel Dengan Faktor Cuaca, Hari Libur dan Hari Besar, Serta Aktivitas Promosi Dengan Model Hybrid Machine Learning.”**

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut,

1. Bagaimana keterkaitan faktor eksternal berupa suhu udara sebagai representasi faktor cuaca, hari libur dan hari besar, serta aktivitas promosi terhadap fluktuasi penjualan ritel mingguan?
2. Bagaimana model *hybrid machine learning* yang dibangun dari kombinasi *Prophet*, *XGBoost*, dan *LSTM* dapat dikembangkan untuk memprediksi penjualan ritel mingguan berdasarkan faktor eksternal tersebut?
3. Sejauh mana tingkat akurasi model *hybrid* yang dikembangkan dibandingkan dengan model tunggal berdasarkan metrik evaluasi *MAE*, *RMSE*, *sMAPE*, dan R^2 ?

1.3 Tujuan

Tujuan penelitian ini sebagai berikut,

1. Menganalisis keterkaitan faktor eksternal berupa suhu udara sebagai representasi faktor cuaca, hari libur dan hari besar, serta aktivitas promosi terhadap pola penjualan mingguan ritel.
2. Mengembangkan model *hybrid machine learning* dengan menggunakan tiga model dasar yaitu, *XGBoost*, *LSTM*, dan *Prophet* yang dibangun dari kombinasi model-model tersebut dengan mempertimbangkan faktor-faktor eksternal.
3. Mengevaluasi performa model *hybrid* dibandingkan dengan model tunggal menggunakan metrik *MAE*, *RMSE*, *sMAPE*, dan R^2

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut,

1. Dengan memanfaatkan kombinasi model *hybrid Prophet*, *LSTM* dan *XGBoost*, penelitian ini dapat memberikan Gambaran mengenai model yang paling sesuai untuk prediksi penjualan ritel mingguan berdasarkan hasil evaluasi.
2. Melalui penerapan *machine learning*, penelitian ini mendukung identifikasi pola musiman, hari libur dan hari besar, serta aktivitas promosi yang dapat digunakan sebagai bahan pertimbangan dalam penyusunan strategi promosi dan persediaan.
3. Dengan penerapan hasil prediksi sebagai dasar analisis, penelitian ini dapat mendukung pengambilan keputusan berbasis data guna meningkatkan efisiensi operasional pada sektor ritel.

1.5 Ruang Lingkup

Ruang lingkup penelitian ini meliputi batasan-batasan sebagai berikut,

1. **Dataset:** Menggunakan *Walmart Sales Forecast Dataset* dari Kaggle yang memuat variabel *Store*, *Dept*, *Date*, *Weekly_Sales*, *Temperature*, *Fuel_Price*, *CPI*, *Unemployment*, *MarkDown1-5*, *IsHoliday*, *Type*, dan *Size*. Dalam penelitian ini, *Weekly_Sales* digunakan sebagai variabel target, sedangkan *Temperature*, *IsHoliday*, dan *MarkDown1-5* digunakan sebagai variabel utama yang sesuai dengan fokus penelitian. Selain itu, *Fuel_Price*, *CPI*, dan *Unemployment* digunakan sebagai variabel ekonomi pendukung untuk memperkaya informasi model, tetapi tidak menjadi fokus utama analisis.
2. **Pembagian Data:** Data yang digunakan untuk pelatihan dan evaluasi model berasal dari *train.csv* karena memiliki variabel target *Weekly_Sales*. Data tersebut kemudian dibagi menjadi *train*, *validation*, dan *test* internal dengan proporsi 70%, 15%, dan 15%

berdasarkan urutan waktu. Sementara itu, *test.csv* tidak digunakan untuk evaluasi model karena tidak memiliki variabel target *Weekly_Sales*, sehingga digunakan pada tahap *deployment* sebagai prediksi akhir.

3. **Faktor Cuaca:** Faktor cuaca dalam penelitian ini dibatasi secara khusus hanya pada variabel *Temperature* yang digunakan sebagai indikator suhu udara pada wilayah tempat toko beroperasi dalam satuan *Fahrenheit*. Unsur meteorologis lainnya, seperti curah hujan, kelembapan, kecepatan angin, atau kondisi cuaca ekstrem, tidak dimasukkan ke dalam cakupan analisis penelitian.
4. **Faktor Hari Libur dan Hari Besar:** Faktor hari libur dan hari besar direpresentasikan oleh variabel *IsHoliday*. Variabel ini digunakan untuk membedakan minggu yang termasuk periode hari libur atau hari besar dan minggu biasa.
5. **Faktor Aktivitas Promosi:** Faktor aktivitas promosi direpresentasikan oleh variabel *Markdown1–5*. Variabel tersebut digunakan untuk menunjukkan adanya aktivitas promosi atau potongan harga yang berkaitan penjualan mingguan. Dalam proses pemodelan, *Markdown1–5* juga dapat diolah menjadi fitur turunan seperti *Total_Markdown*, *Log_Total_Markdown*, *IsPromo*, dan *Markdown_Active_Count*.
6. **Fokus Penelitian:** Menganalisis keterkaitan faktor eksternal yaitu suhu udara, hari libur dan hari besar, serta aktivitas promosi terhadap penjualan mingguan toko ritel. Keterkaitan faktor-faktor tersebut dianalisis pada tingkat mingguan sesuai struktur variabel target yang digunakan, sehingga hasil penelitian tidak dimaknai sebagai bukti kausalitas langsung, melainkan sebagai hubungan yang ditunjukkan melalui eksplorasi data dan pemodelan prediktif.
7. **Model:** Menggunakan model yang terdiri atas tiga model tunggal yaitu *Prophet*, *XGBoost*, dan *LSTM* serta model *hybrid* yang dibangun berdasarkan kombinasi dari ketiga model tersebut. kinerja model *hybrid* selanjutnya dibandingkan dengan kinerja masing-masing model tunggal.
8. **Hyperparameter:** Optimasi dilakukan melalui *grid search* untuk memperoleh kombinasi parameter terbaik secara sistematis.
9. **Tools:** Implementasi dilakukan menggunakan bahasa pemrograman *python* di platform *Google Colab*.
10. **Evaluasi:** Metrik yang digunakan meliputi *MAE*, *RMSE*, *sMAPE*, dan R^2 untuk mengevaluasi performa prediksi model.