

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Toko retail, termasuk minimarket dan grocery store, merupakan bagian integral dari sistem distribusi barang di masyarakat. Toko-toko ini menyediakan berbagai kebutuhan sehari-hari mulai dari bahan makanan, produk kebersihan, hingga kebutuhan rumah tangga lainnya [1]. Keberadaan toko retail tidak hanya menjangkau kota-kota besar, tetapi juga telah merambah ke daerah pinggiran dan pedesaan, menjadikannya salah satu sektor bisnis yang paling mudah diakses oleh masyarakat dari berbagai kalangan [2].

Fenomena maraknya toko retail di berbagai tempat tidak terlepas dari kemudahan akses yang ditawarkan serta kecenderungan masyarakat modern yang menginginkan efisiensi dalam berbelanja [3]. Pola konsumsi masyarakat yang bersifat praktis dan instan semakin mendorong pertumbuhan toko-toko retail dalam skala besar maupun kecil [4]. Minimarket yang buka hingga malam hari bahkan 24 jam, menjadi pilihan utama bagi masyarakat urban yang memiliki mobilitas tinggi [5].

Kebutuhan konsumen terhadap produk-produk harian maupun bulanan menjadi faktor utama dalam keberlangsungan bisnis retail [6]. Tingkat konsumsi masyarakat yang cenderung tinggi, baik karena faktor kebutuhan rutin maupun gaya hidup yang menjadi peluang sekaligus tantangan bagi pelaku bisnis retail dalam menjaga ketersediaan barang secara konsisten [7]. Tingkat konsumsi yang fluktuatif, terutama dipengaruhi oleh momen-momen khusus seperti awal bulan, akhir pekan, dan hari besar keagamaan, menuntut pelaku usaha untuk dapat menyesuaikan penawaran dengan permintaan secara presisi [8].

Permintaan pasar yang dinamis menjadi tantangan tersendiri. Misalnya, menjelang bulan Ramadhan atau Hari Raya, permintaan terhadap produk tertentu seperti bahan makanan pokok dan barang kebutuhan rumah tangga meningkat tajam. Sebaliknya, pada bulan-bulan tertentu, terjadi penurunan penjualan yang signifikan [9]. Fluktuasi ini berkaitan erat dengan tren penjualan yang bersifat musiman dan tahunan. Oleh karena itu, pemahaman terhadap tren penjualan sangat penting bagi pelaku bisnis dalam menyusun strategi logistik dan pemasaran.

Permasalahan muncul ketika pelaku bisnis retail tidak memiliki sistem yang

baik untuk mengantisipasi fluktuasi permintaan tersebut. Ketidaktepatan dalam memprediksi permintaan barang dapat menimbulkan dua masalah utama: overstock dan understock [10]. Overstock menyebabkan biaya penyimpanan meningkat serta risiko kerugian akibat produk yang tidak laku. Sebaliknya, understock menyebabkan kehilangan potensi penjualan dan menurunnya kepuasan pelanggan [11, 12]. Kedua kondisi ini sangat merugikan, terutama bagi usaha kecil dan menengah yang memiliki keterbatasan modal dan sumber daya.

Dalam konteks inilah, prediksi penjualan menjadi komponen yang sangat penting. Kemampuan untuk memprediksi penjualan secara akurat memungkinkan toko retail untuk mengelola inventaris dengan lebih efisien [13]. Prediksi yang baik membantu dalam menyusun rencana pengadaan barang, distribusi stok, serta strategi promosi yang tepat sasaran. Namun, melakukan prediksi penjualan secara manual atau berdasarkan intuisi sering kali tidak memberikan hasil yang memadai, terutama dalam skala data yang besar dan kompleks [14].

Untuk menjawab tantangan tersebut, pendekatan berbasis data atau data-driven menjadi solusi yang relevan. Pendekatan yang digunakan dalam penelitian ini adalah dengan menerapkan algoritma Random Forest Regressor [15]. Random Forest adalah metode ensemble learning berbasis pohon keputusan yang menggabungkan hasil dari banyak pohon (decision trees) untuk menghasilkan prediksi yang lebih stabil dan akurat [16].

Secara umum, Random Forest bekerja dengan membangun sejumlah decision tree pada subset acak dari data pelatihan dan menggabungkan hasilnya melalui proses voting atau averaging untuk meningkatkan akurasi dan mengurangi risiko overfitting [17]. Dalam konteks prediksi penjualan, algoritma ini dapat mempertimbangkan berbagai faktor seperti waktu, kategori produk, dan tren penjualan untuk menghasilkan prediksi yang presisi. Model ini juga cukup fleksibel dalam menangani data dengan noise dan memiliki performa yang baik pada data tabular seperti data penjualan historis [18].

Studi sebelumnya membahas penerapan teknik data mining untuk peramalan penjualan di sektor retail, yang merupakan tugas krusial dalam menentukan kesuksesan jangka panjang suatu organisasi. Penelitian ini melakukan analisis komparatif terhadap beberapa teknik *supervised learning*, seperti *Multiple Linear Regression*, *Random Forest Regression*, *K-Nearest Neighbors (K-NN)*, *Support Vector Machine (SVM)*, dan *Extra Tree Regression*, untuk membangun model prediksi penjualan yang akurat pada 45 outlet Walmart yang terletak di berbagai lokasi geografis. Faktor-faktor yang mempengaruhi prediksi penjualan meliputi

data penjualan sebelumnya, acara promosi, libur nasional, suhu, harga bahan bakar, Indeks Harga Konsumen (CPI), dan tingkat pengangguran di negara bagian tersebut. Hasil penelitian menunjukkan bahwa model Random Forest Regression memiliki performa kedua terbaik dengan nilai R-squared tertinggi sebesar 98,45%, menunjukkan akurasi prediksi yang sangat tinggi [19].

Studi lain juga menunjukkan bahwa Random Forest adalah salah satu algoritma yang efektif dalam melakukan prediksi penjualan. Penelitian berjudul “Improved Demand Forecasting of a Retail Store Using a Hybrid Machine Learning Model” oleh Dholakia dkk. (2023) membahas pengembangan sistem prediksi permintaan ritel dengan pendekatan model hibrida berbasis machine learning untuk meningkatkan akurasi peramalan. Studi ini menggunakan data historis penjualan harian dari sebuah toko ritel yang mencakup 1.000 SKU (Stock Keeping Unit), dengan variabel seperti harga produk, harga pesaing, dan ketersediaan stok. Beberapa algoritma regresi dibandingkan dalam penelitian ini, termasuk Random Forest, Linear Regression, Polynomial Regression, Decision Tree Regression, dan Support Vector Regression. Hasil evaluasi menunjukkan bahwa Random Forest memberikan performa terbaik sebagai model tunggal dengan nilai MAPE (Mean Absolute Percentage Error) sebesar 8,0%. Untuk mengoptimalkan hasil, penulis merancang model hibrida yang menggabungkan dua model terbaik (random forest dan linear regression) berdasarkan performa tiap SKU dengan pendekatan rata-rata tertimbang. Model gabungan ini mampu menurunkan MAPE (Mean Absolute Percentage Error) menjadi 7,76%, menunjukkan peningkatan akurasi secara signifikan. Penelitian ini menyimpulkan bahwa model hibrida efektif dalam meningkatkan prediksi, namun random forest tetap disimpulkan sebagai model tunggal dengan performa terbaik. [20]. Hal ini semakin memperkuat keputusan untuk menggunakan Random Forest dalam proyek ini.

Pada jurnal *Advanced Predictive Model for Optimizing Inventory Management and Demand Forecasting in Smart Logistic* karya Mashael M. Khayyat dan Dr. Shashi Kant Gupta, peneliti mengembangkan model prediksi lanjutan untuk meningkatkan akurasi permintaan dan efisiensi manajemen persediaan dalam sistem logistik cerdas. Mereka menggunakan dataset dari toko ritel yang merepresentasikan data historis penjualan dan stok toko ritel, dengan atribut seperti Units Ordered, Price, Discount, Inventory Level, dan faktor eksternal lainnya. Model yang diuji mencakup machine learning dan big data analytics, serta dievaluasi menggunakan metrik seperti MAE (Mean Absolute Error) dan RMSE (Root Mean Square Error) untuk menilai akurasi prediksi permintaan dan

manajemen inventori. Hasil eksperimen menunjukkan peningkatan signifikan performa sistem prediksi dibanding metode tradisional: RMSE menurun hingga 10–20%, dan MAE juga menunjukkan perbaikan, menandakan kemampuan model dalam mengurangi kesalahan prediksi serta meningkatkan efisiensi pengaturan stok. Pendekatan ini mencerminkan kemajuan dalam integrasi data real-time dan algoritma cerdas untuk optimasi rantai pasok modern. [21]

Berdasarkan permasalahan dan pembahasan di atas, penelitian ini berfokus pada penerapan algoritma Random Forest Regressor untuk menyelesaikan permasalahan prediksi penjualan dalam sektor retail. Dengan memanfaatkan data historis penjualan dan teknologi machine learning, model yang dikembangkan mampu membantu pelaku usaha dalam mengelola stok barang secara lebih efisien, menghindari risiko overstock maupun understock, serta memberikan wawasan strategis yang bermanfaat dalam pengambilan keputusan bisnis.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma Random Forest Regressor dapat memprediksi penjualan produk?
2. Berapa nilai RMSE (Root Mean Square Error), MAE (Mean Absolute Error), dan MAPE (Mean Absolute Percentage Error) dari model random forest regressor dalam memprediksi penjualan produk?

1.3 Batasan Permasalahan

Untuk menjaga fokus dan ruang lingkup penelitian, berikut adalah batasan permasalahan yang ditetapkan dalam penelitian ini:

1. Fokus penelitian akan berada pada seberapa baik performa algoritma Random Forest dalam memprediksi penjualan produk dengan metrik evaluasi MAE (Mean Absolute Error), RMSE (Root Mean Square Error), dan MAPE (Mean Absolute Percentage Error).
2. Pembahasan dan diskusi akan difokuskan pada seberapa berpengaruh kombinasi hyperparameter terhadap kinerja model, hyperparameter

yang digunakan adalah `n_estimators`, `max_depth`, `min_samples_leaf`, dan `min_samples_split`.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma *Random Forest* pada model prediksi penjualan produk toko retail guna melihat efektivitas algoritma terhadap model berdasarkan dataset yang digunakan
2. Mengukur RMSE (Root Mean Square Error), MAE (Mean Absolute Error), dan MAPE (Mean Absolute Percentage Error) dari model prediksi penjualan produk yang dikembangkan untuk menentukan seberapa efektif algoritma *Random Forest* dalam prediksi penjualan produk toko retail berdasarkan dataset yang digunakan

1.5 Manfaat Penelitian

Manfaat dari penelitian ini memberikan salah satu ide atau acuan referensi yang dapat disesuaikan kembali berdasarkan dataset yang berlaku, bagi pelaku usaha retail dalam meningkatkan efisiensi pengelolaan stok barang serta mengurangi kerugian melalui prediksi penjualan dengan machine learning.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

1. Bab 1 PENDAHULUAN
Bab ini membahas latar belakang penelitian, rumusan masalah, batasan masalah, tujuan, dan manfaat penelitian yang dilakukan
2. Bab 2 LANDASAN TEORI
Bab ini menguraikan teori-teori yang mendasari penelitian ini, seperti prediksi penjualan toko retail, machine learning, supervised learning, interquartile range, transformasi logaritmik, deployment, random forest, dan metrik evaluasi yang digunakan dalam penelitian.

3. Bab 3 METODOLOGI PENELITIAN

Metodologi penelitian berisi penjelasan mengenai tahapan-tahapan yang ditempuh selama proses penelitian. Pada bagian ini disajikan flowchart yang menggambarkan jalannya penelitian serta algoritma yang diterapkan dalam studi ini.

4. Bab 4 HASIL DAN DISKUSI

Bab Hasil dan Diskusi memaparkan temuan dari penelitian beserta analisis yang dilakukan terhadapnya. Penyajian hasil disertai dengan visualisasi berupa gambar, tabel, serta uraian tertulis yang menjelaskan secara rinci pencapaian dan interpretasi dari penelitian ini.

5. Bab 5 KESIMPULAN DAN SARAN

Bagian ini berisi kesimpulan dari hasil penelitian yang telah dilaksanakan serta saran yang dapat diberikan untuk pengembangan atau penelitian selanjutnya.

