

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam era digital yang berkembang pesat, analisis data penjualan telah menjadi komponen inti bagi perusahaan untuk memahami perilaku konsumen dan preferensi pasar. Pemahaman yang mendalam terhadap pola konsumsi ini memungkinkan perusahaan untuk mengidentifikasi produk yang memiliki performa penjualan tinggi dan rendah, sehingga mereka dapat mengembangkan strategi pemasaran yang lebih terarah dan efektif. Pada sektor ritel, kemampuan untuk mengklasifikasikan volume penjualan berdasarkan kategori produk dapat membantu bisnis dalam mengalokasikan sumber daya dan menentukan prioritas pemasaran yang lebih efisien. Penggunaan *Machine Learning* dalam memprediksi penjualan diharapkan dapat meningkatkan akurasi dan memberikan pandangan yang lebih tajam mengenai tren pasar (Riza, 2022). Dalam konteks industri ritel, perkembangan yang pesat bahkan melampaui pertumbuhan ekonomi Indonesia. Berdasarkan data Aprindo (Asosiasi Perusahaan Ritel Indonesia), pertumbuhan ritel di tahun 2018 mencapai 9-10%, naik signifikan dibandingkan dengan tahun 2017 yang hanya kurang dari 7% (Novianti & Elisa, 2020).

Minimarket merupakan salah satu bentuk usaha ritel modern yang telah mengalami pertumbuhan sangat pesat, yakni lebih dari 15% per tahun. Minimarket ini telah menjadi salah satu saluran penting dalam memenuhi kebutuhan konsumen di tengah persaingan yang ketat, di mana bisnis ritel seperti minimarket X di Indonesia memanfaatkan teknologi canggih untuk memproses data dan memprediksi angka penjualan produk secara instan (Siregar, 2023). Salah satu metode yang umum digunakan untuk prediksi penjualan dan pengelolaan persediaan adalah algoritma *Random Forest*, yang terbukti mampu meningkatkan akurasi prediksi penjualan serta optimalisasi pengelolaan persediaan dan pengambilan keputusan bisnis (Zyen, 2024). Akurasi prediksi yang tinggi memungkinkan pengelola bisnis untuk mengatur produksi dan persediaan dengan lebih baik, sehingga mampu meningkatkan kepuasan pelanggan melalui ketersediaan produk yang lebih konsisten (Fattah *et al.*, 2022). Dalam

penelitian ini, cara yang digunakan untuk menentukan dari data yang sudah ada yaitu dengan menggunakan metode algoritma *Random Forest*. Algoritma *Random Forest* menjadi metode yang efektif dalam klasifikasi dengan mengurangi varians dan meningkatkan stabilitas model melalui teknik bootstrapping (Kurniawan *et al.*, 2023).

Klasifikasi produk berdasarkan dari platform penjualan sangat penting untuk meningkatkan efisiensi pemasaran, di mana *Random Forest* sebagai metode yang berperan sebagai teknik klasifikasi. Cara yang digunakan dalam *Random Forest* ini yaitu dengan menggunakan metode gabungan pohon (ensemble tree), *Random Forest* meningkatkan akurasi klasifikasi dengan mengakumulasi hasil dari setiap pohon keputusan yang dibentuk secara acak (Rindiyan *et al.*, 2022). Selain keunggulannya dalam akurasi, *Random Forest* juga memberikan fleksibilitas bagi pengguna untuk menyesuaikan jumlah pohon yang dibentuk, yang dapat dioptimalkan sesuai dengan kebutuhan dataset dan tujuan analisis. Dengan kinerja yang konsisten dan akurasi yang tinggi, *Random Forest* telah terbukti menjadi metode yang andal untuk meningkatkan efisiensi dan ketepatan klasifikasi produk di berbagai platform penjualan. Dalam beberapa penelitian, performa *Random Forest* menunjukkan keunggulan dibandingkan metode lain, seperti Naïve Bayes yang hanya mencapai akurasi 55.80% dibandingkan 70.60% pada kasus klasifikasi penyakit liver (Lubis *et al.*, 2022).

Visualisasi interaktif memainkan peran krusial dalam analisis data, terutama ketika data memiliki banyak variabel dan menghasilkan informasi yang memerlukan interpretasi mendalam. Dengan visualisasi interaktif, pengguna dapat mengeksplorasi data lebih dalam, mengidentifikasi pola yang tidak terlihat melalui metode statis, dan memperoleh wawasan yang lebih akurat (Irwan *et al.*, 2024). Visualisasi interaktif dapat meningkatkan pemahaman terhadap hasil model dan mempercepat pengambilan keputusan dalam bisnis. Selain itu, Visualisasi interaktif memungkinkan pengguna untuk berinteraksi langsung dengan data, memberikan fleksibilitas dalam menyaring, memperbesar, atau memfokuskan area tertentu untuk analisis yang lebih mendalam. Selain itu, pendekatan ini meningkatkan kolaborasi antar pemangku kepentingan, karena hasil yang disajikan secara visual lebih mudah dipahami dan dianalisis bersama-sama dalam proses pengambilan keputusan (Hilyah, 2024).

Penelitian sebelumnya yang dilakukan oleh (Lubis *et al.*, 2022) menyatakan hasil yang diperoleh saat penelitian ini dengan menggunakan Indian Liver Patient Dataset (ILPD) diketahui dari hasil penelitian yaitu dengan metode Naïve Bayes diperoleh nilai akurasi 55.80% dan metode *Random Forest* memperoleh nilai akurasi 70.60%. Dapat disimpulkan bahwa Naïve Bayes dan *Random Forest* memiliki performa kinerja yang berbeda di dalam proses perhitungan klasifikasi pada kasus penyakit liver yang dimana *Random Forest* lebih unggul peroleh akurasi yang dihasilkan. Untuk meningkatkan penelitian di masa depan, disarankan menggunakan lebih banyak metode klasifikasi, memperluas dataset, menerapkan validasi silang yang lebih ketat, dan menganalisis fitur-fitur yang paling berpengaruh. Langkah-langkah ini diharapkan dapat meningkatkan akurasi, generalisasi, dan interpretabilitas model dalam klasifikasi.

Menurut penelitian yang dilakukan oleh (Utomo, 2024) yang menggunakan algoritma *Random Forest* dan teknik Boosting dalam sistem rekomendasi produk berbasis ulasan pengguna e-commerce menunjukkan hasil yang signifikan dalam meningkatkan akurasi klasifikasi. Dalam sebuah penelitian, penggabungan *Random Forest* dengan *XGBoost* menghasilkan nilai akurasi tertinggi sebesar 87,25% yang lebih baik dibandingkan dengan kombinasi metode lain seperti Gradient Boosting dan LightGBM. Penelitian tersebut membuktikan bahwa teknik ensemble *Random Forest-XGBoost* mampu memberikan akurasi dan stabilitas yang lebih tinggi dalam pemodelan data yang kompleks, seperti ulasan pengguna produk e-commerce. Hasil ini menunjukkan bahwa kombinasi algoritma ensemble dapat memberikan solusi yang lebih akurat dan dapat diandalkan dalam menangani data berukuran besar dan beragam. Selain itu, penelitian tersebut juga menggunakan evaluasi presisi, recall, dan F1-Score untuk menilai performa model, sehingga memberikan pemahaman yang komprehensif mengenai efektivitas algoritma dalam klasifikasi data.

Pada temuan sebelumnya yang dilakukan oleh (Ambarwati, 2024) dari hasil data set prediksi, hasil prediksi penjualan CV Analys Jaya Computindo dengan *Neural Network* di RapidMiner menunjukkan efektivitas dan akurasi tinggi, terlihat dari perbedaan kecil antara nilai aktual dan prediksi. Tingkat akurasi ini tercermin dari nilai

RMSE yang rendah, yaitu 12,834 untuk segmen Toko, 1,895 untuk segmen Perusahaan, dan 0,385 untuk segmen perorangan. Analisis ini memberikan rekomendasi strategi bauran pemasaran yang diharapkan membantu manajemen dalam menentukan target penjualan masa depan. Penelitian sebelumnya menunjukkan beberapa keterbatasan, terutama dalam cakupan data yang hanya mencakup segmen toko, perusahaan, dan perorangan, yang dapat mengurangi generalisasi hasil. Meski algoritma *Neural Network* terbukti efektif sebagai model prediktif dengan tingkat akurasi tinggi, model ini kurang optimal untuk klasifikasi data kompleks dan memerlukan pemahaman mendalam terkait pemilihan parameter.

Penelitian ini berfokus pada penggunaan Algoritma *Random Forest* untuk mengklasifikasikan volume dan memvisualisasikan penjualan produk ritel berdasarkan kategori produk dan kebiasaan pembelian konsumen. Algoritma ini akan digunakan untuk menilai suatu data dari penjualan sebelumnya. Selain itu, algoritma tersebut juga digunakan untuk mengidentifikasi kriteria utama yaitu kategori produk, quantity, dan total. Kemudian, untuk membangun model prediksi yang akurat. Perusahaan ritel diharapkan mengoptimalkan manajemen stok, meningkatkan teknik promosi, dan meningkatkan distribusi serta pengambilan keputusan harga.

Berdasarkan permasalahan diatas, peneliti tertarik untuk melakukan penelitian dengan topik "Prediksi Volume Penjualan Ritel Berdasarkan Kategori Produk Menggunakan Algoritma *Random Forest* dan *XGBoost*" Implementasi ini bertujuan untuk meningkatkan akurasi serta prediksi data penjualan, mengklasifikasikan kategori penjualan, dan memvisualisasikan data dari hasil prediksi, sehingga mendorong pertumbuhan perusahaan yang lebih besar dan responsif terhadap pasar.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini, diantaranya :

1. Kurangnya akurasi dalam memprediksi data penjualan produk ritel ketika menggunakan algoritma lain seperti *XGBoost*, yang menyebabkan hasil prediksi kurang optimal untuk kebutuhan bisnis.
2. Belum diterapkannya teknik klasifikasi berbasis algoritma *Random Forest* untuk mengkategorikan produk dalam analisis penjualan, yang dapat membantu meningkatkan pengelolaan kategori produk.
3. Keterbatasan visualisasi dalam pengolahan data penjualan, yang menyulitkan perusahaan dalam menganalisis pola dan tren penjualan produk.

## 1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk menjawab beberapa pertanyaan utama, yaitu :

1. Membantu akurasi prediksi penjualan produk ritel berdasarkan kategori pembelian dengan menerapkan perbandingan algoritma *Random Forest* dan algoritma *XGBoost*.
2. Meningkatkan sistem klasifikasi yang dapat membantu perusahaan dalam pengelolaan stok berdasarkan hasil prediksi permintaan produk.
3. Memvisualisasikan data penjualan produk agar mempermudah untuk di analisis dengan menggunakan *Machine Learning*.

## 1.4 Batasan Masalah

1. Penelitian ini menggunakan algoritma *Random Forest* dan membandingkan dengan algoritma *XGBoost* untuk mengelola data ritel yang beragam.
2. Prediksi volume penjualan berdasarkan kategori produk (Buah-buahan, Dry Food, Ikan Segar, Non Food, dan Sayur-Sayuran) dan berapa banyak penjualan dalam satu bulan agar bulan selanjutnya dapat terprediksi.
3. Sistem ini dilakukan dengan pengujian dataset Penjualan ritel sales di XYZ dalam bulan tertentu untuk membantu pengolahan prediksi.

## 1.5 Manfaat Penelitian

Manfaat penelitian dengan judul “Prediksi Volume Penjualan Ritel Berdasarkan Kategori Produk Menggunakan Algoritma *Random Forest* dan *Xgboost*” mencakup berbagai aspek yang akan mendapatkan dampaknya, yaitu :

1. Masyarakat umum, khususnya pelanggan dengan mengembangkan strategi penjualan yang lebih efisien dan tepat sasaran dari perusahaan ritel. Dengan prediksi penjualan yang lebih akurat, konsumen akan lebih mudah menemukan produk yang mereka butuhkan tanpa harus menghadapi kekurangan stok atau ketidaktersediaan barang pada ritel.
2. Data analyst, yang bekerja di sektor ritel atau yang tertarik untuk mengembangkan model prediksi dan klasifikasi penjualan ritel berbasis *Machine Learning*, sehingga mereka dapat membuat model yang lebih akurat dan stabil dalam pengelolaan data ritel.
3. Perusahaan ritel, memberikan wawasan mengenai variabel-variabel yang paling berpengaruh terhadap volume penjualan, yang berpotensi meningkatkan efisiensi manajemen stok, pemasaran, penetapan harga, serta profitabilitas secara keseluruhan.

