PREDIKSI PENJUALAN WARUNG KOPI OI MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST

Yulita Marilaeta Nurak, Syahroni Wahyu Iriananda, Fitri Marisa
Teknik Informatika; Universitas Widyagama Malang
Jl. Borobudur No.35, Mojolangu, Kec. Lowokwaru, Kota Malang, Indonesia
yulitanurak@gmail.com

ABSTRAK

Pesatnya perkembangan bisnis kedai kopi mendorong para pemilik usaha untuk lebih cermat dalam mengelola stok bahan baku, guna mengantisipasi permintaan pelanggan. Warung Kopi OI menghadapi kendala dalam pengelolaan persediaan yang selama ini dilakukan secara manual, sering kali menyebabkan pembelian bahan baku yang tidak terencana. Penelitian ini bertujuan untuk memprediksi penjualan menggunakan dua algoritma pembelajaran mesin, yakni Random Forest dan XGBoost. Data yang dianalisis diperoleh dari Kaggle, mencakup transaksi penjualan selama enam bulan. Proses penelitian meliputi pengumpulan data, tahap pra-pengolahan, implementasi algoritma, serta evaluasi model berdasarkan akurasi, RMSE, dan MAPE. Hasil penelitian menunjukkan bahwa penggunaan rasio data latih dan uji sebesar 80:20 memberikan hasil paling optimal dengan akurasi tertinggi, yakni 72%, pada kedua metode. Secara keseluruhan, Random Forest terbukti lebih stabil dibandingkan XGBoost dengan rata-rata akurasi masing-masing 64% dan 62%. Temuan ini menegaskan bahwa pemilihan rasio data yang tepat berpengaruh signifikan terhadap performa model prediksi. Penelitian ini diharapkan dapat memberikan rekomendasi kepada Warung Kopi OI dalam pengelolaan bahan baku secara lebih efisien, sehingga mampu mengurangi kelebihan maupun kekurangan stok.

Kata kunci : *peramalan penjualan, Random Forest, XGBoost, rasio data*

1. PENDAHULUAN

Indonesia memiliki peran yang signifikan dalam produksi dan konsumsi kopi dunia, menempati posisi keempat sebagai negara produsen terbesar. Selain itu, Indonesia juga masuk dalam sepuluh besar negara konsumen kopi global, menduduki urutan ketujuh. Selama periode 2017 hingga 2018, terjadi lonjakan konsumsi kopi sebesar 13,84%, setara dengan 314 ton berbagai jenis kopi. Peningkatan konsumsi kopi di Indonesia terus berlangsung sejak 2016 hingga 2018 [1].

Perkembangan pesat bisnis kedai kopi mengakibatkan persaingan yang semakin ketat. Pemilik kedai berlomba-lomba memberikan pelayanan terbaik bagi pelanggan, misalnya dengan menyediakan fasilitas yang menarik, seperti ruangan yang bersih dan nyaman serta suasana yang cozy. Beberapa kedai kopi bahkan menawarkan layanan tambahan seperti lounge, Wi-Fi gratis, AC, hingga desain interior yang menarik [2].

Warung Kopi OI merupakan bisnis kedai kopi yang berlokasi di Jl. Ciujung No.8, Purwantoro, Kec. Blimbing, Kota Malang. Tantangan utama yang dihadapi Warung Kopi OI adalah kurangnya sistem kontrol yang efektif dalam penyediaan bahan baku. Hingga saat ini, pemilik bisnis masih mengandalkan perkiraan pribadi tanpa menggunakan metode yang terstruktur untuk menentukan jumlah bahan baku yang harus dipesan. Akibatnya, pemesanan bahan baku sering kali dilakukan secara mendadak ketika stok hampir habis. Pemesanan mendadak ini kadang dilakukan ke pemasok lain karena pemasok tetap sudah memiliki jadwal pengiriman tersendiri, sehingga kualitas bahan baku menjadi tidak konsisten dan berdampak pada cita rasa kopi yang disajikan.

Tanpa pengelolaan stok yang akurat, risiko terjadinya kelebihan atau kekurangan bahan baku menjadi lebih besar. Ketika stok berlebih, biaya penyimpanan akan meningkat, sementara jika stok habis saat permintaan tinggi, warung tidak bisa memenuhi kebutuhan pelanggan tepat waktu, yang berpotensi menurunkan kepercayaan pelanggan [3].

Kelebihan stok (*overstock*) juga menimbulkan pemborosan karena biaya penyimpanan dan pemeliharaan yang tinggi [4].

Oleh sebab itu, penentuan persediaan yang efisien sangat penting untuk menjaga kelancaran proses produksi [5].

Salah satu solusi yang dapat diterapkan untuk mengatasi masalah ini adalah menggunakan metode prediksi penjualan, sehingga jumlah bahan baku yang dipesan dapat disesuaikan dengan kebutuhan.

Penelitian yang dilakukan oleh Rindiyani menggunakan metode Random Forest untuk memprediksi penjualan produk di Omah Branded. Hasil penelitian menunjukkan tingkat akurasi sebesar 92% berdasarkan perhitungan confusion matrix, dengan produk terlaris adalah kaos [6].

Penelitian Alhamdi menggunakan metode XGBoost untuk meramalkan kebutuhan obat di Rumah Sakit OI. Permasalahan yang dihadapi adalah perencanaan kebutuhan obat masih dilakukan secara manual sehingga menghambat proses perencanaan. Dalam penelitian ini, data penggunaan obat vital dan esensial dari tahun 2017 hingga 2022 digunakan sebagai data utama, sementara data cuaca sebagai fitur eksternal. Hasilnya menunjukkan bahwa model

XGBoost untuk obat vital memiliki rata-rata skor RMSE sebesar 17,34 dan MAE sebesar 13,03. Namun, masih ada peluang untuk menurunkan tingkat kesalahan lebih lanjut. Untuk obat esensial, model ini perlu ditingkatkan agar mencapai performa yang lebih optimal [7].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan oleh Febrian tentang prediksi penjualan suku cadang di PT Terus Jaya Sentosa Motor menggunakan algoritma Random Forest. Pada penelitian ini, digunakan dataset yang berisi 3.528 data dengan pendekatan metodologi CRISP-DM. Hasil pengujian model menunjukkan bahwa nilai *Root Mean Squared Error* (RMSE) adalah 23,38, *Mean Absolute Error* (MAE) sebesar 13,97, dan *Mean Squared Error* (MSE) mencapai 546,49. Meskipun nilai MSE tergolong cukup besar, model ini tetap dianggap memiliki kinerja yang baik karena nilai RMSE dan MAE yang relatif kecil. Model prediksi yang dikembangkan ini diintegrasikan ke dalam sebuah situs web agar pengguna dapat dengan mudah memprediksi penjualan suku cadang motor di PT Terus Jaya Sentosa Motor [8].

Penelitian yang dilakukan oleh Efendi tentang penerapan algoritma random forest untuk melakukan prediksi terhadap penjualan dan manajemen stok produk Bolen Crispy di Desa Pekalongan. Model ini diterapkan dengan memanfaatkan data penjualan historis guna memperkirakan permintaan produk di masa mendatang. Pengujian dilakukan menggunakan data penjualan selama satu tahun, dengan alokasi 80% untuk pelatihan dan 20% untuk pengujian. Hasil penelitian awal menunjukkan bahwa penerapan Random Forest mampu meningkatkan akurasi prediksi hingga 85% dibandingkan metode konvensional. Dengan prediksi yang lebih akurat, pengelolaan stok menjadi lebih efisien sehingga dapat meminimalkan risiko kekurangan maupun kelebihan persediaan [9].

Penelitian yang dilakukan oleh Wijaya untuk melakukan optimasi prediksi penjualan pada ritel XYZ menggunakan algoritma XGBoost. Penelitian ini mengajukan metode baru untuk prediksi penjualan dengan menerapkan XGBoost, yang dioptimalkan melalui perbandingan tiga metode optimasi: pencarian acak (*random search*), pencarian grid (*grid search*), dan optimasi Bayesian. Tujuannya adalah untuk mendapatkan analisis perbandingan terbaik dan meningkatkan akurasi prediksi. Kebaruan dari model ini terletak pada penentuan nilai optimal untuk setiap metode optimasi yang diterapkan pada XGBoost. Hasil evaluasi menunjukkan bahwa teknik optimasi *grid search* memberikan performa terbaik pada model XGBoost, dengan peningkatan nilai evaluasi R^2 dari 97,31 menjadi 98,41. Analisis model yang diusulkan ini dapat membantu pemilik bisnis ritel dalam melakukan prediksi penjualan secara lebih akurat untuk mendukung perkembangan proses bisnis [10].

2.2. Random Forest

Random Forest membangun suatu model dengan menggunakan beberapa pohon keputusan (*decision tree*) dan menggabungkan prediksinya untuk mendapatkan prediksi yang lebih akurat dan stabil daripada hanya mengandalkan satu pohon keputusan decision tree. Random Forest merupakan pengembangan lebih lanjut dari CART yang mengaplikasikan *Bootstrap Aggregating* (*Bagging*) dan pemilihan fitur secara acak untuk menghindari *overfitting* pada kumpulan data kecil [11].

Adapun proses pembentukan random forest sebagai berikut [12].

a. Pemilihan Sampel Secara Acak: Sampel data diambil secara acak dengan metode bootstrap dari dataset pelatihan yang tersedia, di mana ukuran sampel yang diambil sama dengan ukuran dataset pelatihan awal.
b. Pemilihan Subset Fitur Acak: Dari sejumlah fitur yang ada, subset fitur dipilih secara acak untuk membangun pohon keputusan. Biasanya, jumlah fitur yang dipilih jauh lebih sedikit dibandingkan dengan total fitur yang tersedia.
c. Proses Pembuatan Pohon Keputusan: Pohon keputusan dibangun menggunakan sampel data dan subset fitur terpilih dengan algoritma seperti ID3, C4.5, atau CART. Pemisahan data dilakukan berdasarkan fitur yang memiliki nilai informasi tertinggi, dengan tujuan mengurangi tingkat variasi dalam simpul pohon.
d. Pembentukan Kumpulan Pohon: Langkah-langkah pemilihan sampel, pemilihan subset fitur, dan pembangunan pohon keputusan diulangi berkali-kali hingga terbentuk sekumpulan pohon keputusan yang terdiri dari beberapa model berbeda.
e. Pengambilan Keputusan Akhir: Hasil prediksi akhir dalam Random Forest diperoleh melalui mekanisme voting mayoritas untuk klasifikasi atau dengan menghitung rata-rata nilai prediksi untuk kasus regresi dari seluruh pohon keputusan dalam ensambel.

2.3. *Extreme Gradient Boosting* (XGBOOST)

Extreme Gradient Boosting atau XGBoost dikembangkan oleh Chen & Guestrin (2016), XGBoost merupakan salah satu metode boosting yaitu kumpulan pohon keputusan yang pembangunan pohon berikutnya akan bergantung pada pohon sebelumnya. Pohon pertama dalam XGBoost akan lemah dalam melakukan klasifikasi dengan inisialisasi probability telah ditentukan dan kemudian akan dilakukan update bobot pada setiap pohon yang dibangun sehingga menghasilkan kumpulan pohon klasifikasi yang kuat [13].

Berikut langkah-langkah utama dalam proses pembentukan model XGBoost [14]:

a. Inisialisasi Model Awal, proses dimulai dengan menetapkan model awal yang sederhana, seperti

memprediksi nilai rata-rata target dari semua data latih.

b. Menghitung Kesalahan (Residual), selisih antara hasil prediksi model awal dan nilai aktual dihitung sebagai residual. Residual ini mencerminkan besarnya kesalahan yang ingin dikurangi pada tahap selanjutnya.
c. Membangun Pohon Keputusan Baru, pohon keputusan dibuat untuk memprediksi residual yang telah dihitung. Pohon ini berfungsi sebagai koreksi terhadap kesalahan yang dihasilkan oleh model sebelumnya.
d. Memperbarui Model, model diperbarui dengan menambahkan prediksi dari pohon baru yang dikalikan dengan faktor pembelajaran (learning rate) untuk mengontrol seberapa besar kontribusi pohon tersebut terhadap model akhir.
e. Proses Iterasi, langkah menghitung residual, membangun pohon keputusan, dan memperbarui model diulang hingga jumlah iterasi yang telah ditentukan tercapai atau model mencapai tingkat akurasi yang diinginkan.
f. Regularisasi, teknik regularisasi diterapkan untuk mengurangi risiko overfitting, sehingga model yang dihasilkan mampu memberikan hasil yang akurat pada data baru.

2.4. Evaluasi Model

Model peramalan yang dikembangkan perlu dievaluasi untuk mengetahui tingkat akurasinya serta menilai kemampuan model dalam memprediksi data dengan tepat. Terdapat berbagai teknik evaluasi peramalan yang sering diterapkan, namun metode yang paling umum digunakan meliputi *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan pengukuran akurasi [15].

2.4.1. *Root Mean Squared Error* (RMSE)

RMSE merupakan salah satu perhitungan yang cukup sering digunakan untuk mengevaluasi peramalan dengan mengukur perbedaan nilai prediksi dengan nilai asli pada data. Semakin kecil nilai RMSE menandakan model yang digunakan untuk melakukan prediksi semakin baik [16].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2 \quad (1)$$

$$RMSE = \sqrt{MSE} \quad (2)$$

2.4.2. *Mean Absolute Percentage Error* (MAPE)

MAPE adalah perhitungan kesalahan rata - rata dari model prediksi secara mutlak, yang di tampilkan dalam bentuk persentase. MAPE pada suatu model rediksi dikatakan sangat baik jika memiliki nilai 10% dan dikatakan buruk jika memiliki nilai diatas 50% [17].

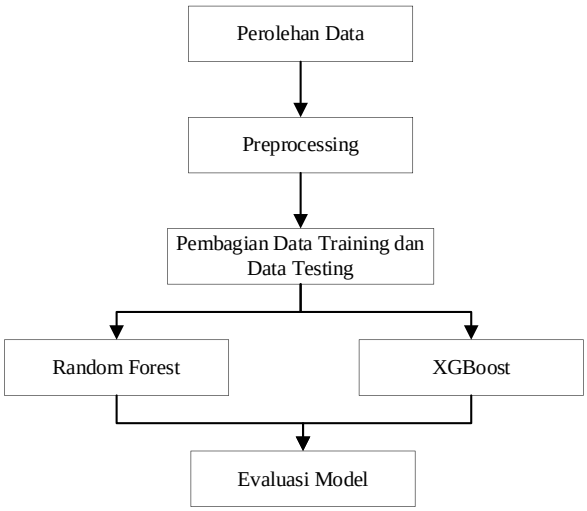
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y-\hat{y}}{y}\right|^2 \quad (3)$$

sedangkan nilai akurasi didapatkan dari 1 dikurangi nilai MAPE.

$$Akurasi = 1 - MAPE \quad (4)$$

3. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis kinerja algoritma random forest dan XGBoost dalam memprediksi penjualan kopi, dengan memanfaatkan dataset yang diambil dari repository Kaggle. Dalam konteks pengembangan strategi pemasaran yang lebih efektif, prediksi penjualan sangat penting untuk meningkatkan efisiensi dan daya saing perusahaan khususnya pengelolaan stok. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Metode penelitian

3.1. Perolehan Data

Data ang digunakan dalam peneltiian ini diperoleh dari website Kaggle.com (https://www.kaggle.com/datasets/keremkarayaz/coffee-shop-sales/data). Data ang dunduh mempunai ekstensi CSV.

3.2. Preprocessing

Pada tahap ini akan melakukan persiapan dataset yang akan digunakan melalui beberapa Langkah, yaitu:

a. Cek data kosong, data yang tidak lengkap dapat mengurangi kualitas informasi yang digunakan oleh model, sehingga terjadi error atau prediksi menjadi kurang akurat, sehingga data yang hilang harus diisi atau dihapus terlebih dahulu untuk menghindari hal tersebut.
b. Cek data kembar, data yang kembar bisa menyebabkan redundansi, di mana informasi yang sama digunakan lebih dari satu kali dalam proses pelatihan model. Hal ini dapat mengarahkan model untuk memberikan bobot lebih pada sampel yang berulang, yang bisa mengakibatkan bias.
c. Seleksi fitur, pada proses ini akan memilih *feature* yang akan digunakan pada proses peramalan, pada peneltian ini adalah tanggal, id produk, dan jumlah penjualan produk.

3.3. Pembagian Data Training dan Data Testing

Data yang telah diproses dibagi menjadi dua bagian untuk memastikan model dapat melakukan generalisasi dengan baik:

a. Data Training, digunakan untuk melatih model machine learning agar dapat mengenali pola dalam data.
b. Data Testing, digunakan untuk menguji performa model dengan data yang belum pernah dilihat sebelumnya.

Pada penelitian ini rasio yang digunakan untuk data training dan data testing 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, dan 10:90.

3.4. Penerapan Random Forest dan XGBoost

Dalam penelitian ini, dua algoritma machine learning diterapkan untuk menganalisis data absensi dan penggajian:

a. Random Forest, parameter utama yang disesuaikan dalam penelitian ini:
   - Jumlah pohon dalam hutan (n_estimators)
   - Kedalaman maksimum pohon (max_depth)
   - Jumlah fitur yang digunakan dalam setiap pembagian (*max_features*)
b. XGBoost, parameter utama yang disesuaikan dalam penelitian ini:
   - Learning rate (eta)
   - Jumlah estimators (n_estimators)
   - Maksimum kedalaman pohon (max_depth)

Model dibangun dengan menggunakan library scikit-learn dan XGBoost dalam Python. Untuk mendapatkan performa terbaik, dilakukan tuning parameter menggunakan *Grid Search*.

3.5. Evaluasi Model

Pada tahap ini hasil pengujian prediksi menggunakan Random Forest dan XGBoost akan dievaluasi menggunakan

a. RMSE, untuk mengukur seberapa besar perbedaan antara nilai aktual dan prediksi.
b. MAPE, untuk mengukur tingkat kesalahan prediksi dalam bentuk persentase.
c. Akurasi. mengukur persentase prediksi yang benar dibandingkan total data uji

4. HASIL DAN PEMBAHASAN

4.1. Perolehan Data

Pada penelitian ini dataset diperoleh dari website kaggle.com (https://www.kaggle.com/datasets/keremkarayaz/coffee-shop-sales/data). Data yang diperoleh dari website kaggle tersebut berformat CSV. Berikut adalah sampel data penjualan warung kopi yang dapat dilihat pada tabel 1.

Tabel 1. Dataset penjualan

Transaction id	Transaction date	…	Product type	Product detail
1	1/1/2023	…	Gourmet brewed coffee	Ethiopia Rg
2	1/1/2023	…	Brewed Chai tea	Spicy Eye Opener Chai Lg
3	1/1/2023	…	Hot chocolate	Dark chocolate Lg
…	…	…	…	…
149115	6/30/2023	…	Barista Espresso	Cappuccino
149116	6/30/2023	…	Regular syrup	Hazelnut syrup

Total record data pada dataset yang digunakan sebanyak 149.456 record.

Tabel 2. Atribut dataset penjualan kopi

No	Atribut Data
0	transaction_id
1	transaction_date
2	transaction_time
3	transaction_qty
4	store_id
5	store_location
6	product_id
7	unit_price
8	product_category
9	product_type
10	product_detail

Ada 11 atribut pada dataset yang ada pada data asli, akan tetapi data tersebut masih akan diolah karena tidak semua atribut akan digunakan.

4.2. Preprocessing

Pada tahap ini akan dilakukan preprocessing data ang terdiri penghapusan data kosong dan data kembar, setelah itu akan dilakukan seleksi atribut yang akan digunakan dalam penelitian ini.

4.2.1. Penghapusan Data Kosong

Proses ini dilakukan untuk menghindari kesalahan pada saat pembuatan model. Berdasarkan hasil pengecekan, tidak ditemukan data kosong dalam dataset ini.

```
transaction_id      0
transaction_date    0
transaction_time    0
transaction_qty     0
store_id            0
store_location      0
product_id          0
unit_price          0
product_category    0
product_type        0
product_detail      0
dtype: int64
transaction_id      0
transaction_date    0
transaction_time    0
transaction_qty     0
store_id            0
store_location      0
product_id          0
unit_price          0
product_category    0
product_type        0
product_detail      0
dtype: int64
```

Gambar 2. Hasil Pengecekan dan Penghapusan Data Kosong

Berdasarkan hasil pengecekan data, tidak ada data kosong pada data penjualan, sehingga tidak ada proses penghapusan data kosong.

4.2.2. Pengahapusan Data Kembar

```
# Cek data duplikat
print(df.duplicated().sum())

# Hapus data duplikat
df = df.drop_duplicates()

# Cek kembali data duplikat setelah dihapus
print(df.duplicated().sum())
```

```
0
0
```

Gambar 3. Hasil Pengecekan dan Penghapusan Data Kosong

Berdasarkan hasil pengecekan data, tidak ada data kembar pada data penjualan, sehingga tidak ada proses penghapusan data kembar.

4.2.3. Seleksi Atribut

Atribut yang digunakan dalam penelitian ini adalah product_id dan transaction_qty. Untuk jumlah transaksi perproduk akan digunakan penjualan dalam satu minggu (total). Dan akan ada penambahan atribut week dikarenakan prediksi penjualan menggunakan data penjualan perminggu, maka pada penelitian ini variabel X diwakili dengan product_id dan week sedangkan variabel Y diwakili dengan total penjualan dalam satu minggu (total). Berikut ini merupakan hasil olah data penjualan perminggu.

	product_id	week	total
0	1	1	5
1	1	2	6
2	1	3	4
3	1	4	2
4	1	5	3
...	...	...	...
2079	87	23	74
2080	87	24	167
2081	87	25	275
2082	87	26	171
2083	87	52	7

[2084 rows x 3 columns]

Gambar 4. Dataset penjualan kopi perminggu

Setelah dilakukan pengolahan data, total record data sebanak 2083.

4.3. Pembagian Data Training dan Data Testing

Pada penelitian ini, rasio data training dan data testing yaitu 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, dan 10:90.

4.4. Penerapan Random Forest dan XGBoost

4.4.1. Penerapan Random Forest

Sebelum melakukan pengujian dengan menggunakan random forest, langkah awal adalah menentukan parameter terbaik yang akan digunakan dalam pengujian.

Tabel 3. Grid search parameter tuning random forest

Parameter	Grid of Values	Parameter Terbaik
n_estimators	100, 200, 300, 400, 500	200
max_depth	None, 10, 20, 30	10
Min_samples_split	2, 5, 10	2
Min_samples_leaf	1, 2, 4	1

Hasil tersebut merupakan nilai parameter optimal yang dapat meningkatkan performa algoritma random forest pada eksperimen yang dilakukan dengan n_estimators 200, max_depth 10, min_samples_split 2, dan min_samples_leaf 1.

4.4.2. Penerapan XGBoost

Sebelum melakukan pengujian dengan menggunakan XGBoost, langkah awal adalah menentukan parameter terbaik yang akan digunakan dalam pengujian.

Tabel 4. Grid search parameter tuning XGBoost

Parameter	Grid of Values	Parameter Terbaik
n_estimators	100, 200, 300, 400, 500	100
max_depth	0.01, 0.05, 0.1, 0.2	0.1
learning_rate	3, 5, 7, 8	5

Hasil tersebut merupakan nilai parameter optimal yang dapat meningkatkan performa algoritma XGBoost pada eksperimen yang dilakukan dengan n_estimators sebesar 100, max_depth 0.1, dan learning rate 5.

4.5. Evaluasi Model

4.5.1. Hasil Pengujian Random Forest

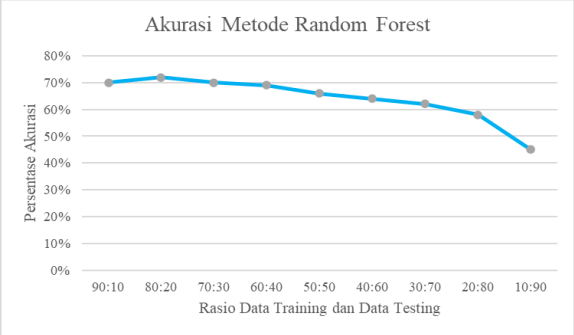
Berikut ini merupakan hasil pengujian menggunakan random forest.

Tabel 5. Hasil pengujian random forest

Rasio Data	RMSE	MAPE	Akurasi
90:10	22.58	30%	70%
80:20	20.55	28%	72%
70:30	21.07	30%	70%
60:40	21.59	31%	69%
50:50	22.03	34%	66%
40:60	24.11	36%	64%
30:70	23.56	38%	62%
20:80	23.71	42%	58%
10:90	25.57	55%	45%
90:10	22.58	30%	70%

Berdasarkan hasil pengujian di atas, rasio data training dan data testing 90:10 menghasilkan RMSE sebesar 22.58, MAPE sebesar 30% dan akurasi sebesar

70%. Rasio 80:20 menghasilkan RMSE sebesar 20.55, MAPE sebesar 28% dan akurasi sebesar 72%. Rasio 70:30 menghasilkan RMSE sebesar 21.07, MAPE sebesar 30% dan akurasi sebesar 70%. Rasio 60:40 menghasilkan RMSE sebesar 21.59, MAPE sebesar 31% dan akurasi sebesar 69%. Rasio 50:50 menghasilkan RMSE sebesar 22.03, MAPE sebesar 34% dan akurasi sebesar 64%. Rasio 40:60 menghasilkan RMSE sebesar 24.11, MAPE sebesar 36% dan akurasi sebesar 64%. Rasio 30:70 menghasilkan RMSE sebesar 23.56, MAPE sebesar 38% dan akurasi sebesar 62%. Rasio 20:80 menghasilkan RMSE sebesar 23.71, MAPE sebesar 42% dan akurasi sebesar 58%. Rasio 10:90 menghasilkan RMSE sebesar 25.57, MAPE sebesar 55% dan akurasi sebesar 45%.



Gambar 5. Grafik hasil pengujian random forest

Berdasarkan grafik pada gambar 5, akurasi tertinggi pada rasio 80:20 sebesar 72% dan akurasi terendah pada rasio 10:90 sebesar 45%. Rasio 80:20 memberikan proporsi yang cukup besar untuk pelatihan model (80%) sambil tetap menyisakan data yang cukup (20%) untuk menguji generalisasi model. Dengan lebih banyak data pelatihan, model dapat lebih baik mempelajari pola dan variasi dalam data. Rasio 80:20 biasanya mengurangi risiko overfitting dibandingkan rasio seperti 90:10. Jika terlalu banyak data digunakan untuk pelatihan, model mungkin belajar terlalu detail tentang data pelatihan dan tidak dapat menggeneralisasi dengan baik pada data baru.

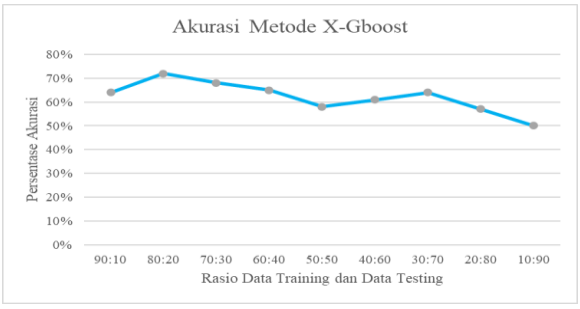
4.5.2. Hasil Pengujian XGBoost

Berikut ini merupakan hasil pengujian menggunakan XGBoost.

Tabel 6. Hasil pengujian XGBoost

Rasio Data	RMSE	MAPE	Akurasi
90:10	22.02	36%	64%
80:20	19.72	29%	72%
70:30	21.87	32%	68%
60:40	23.08	35%	65%
50:50	23.89	42%	58%
40:60	24.44	39%	61%
30:70	25.26	36%	64%
20:80	24.77	43%	57%
10:90	27.66	50%	50%
90:10	22.02	36%	64%

Berdasarkan hasil pengujian di atas, rasio data training dan data testing 90:10 menghasilkan RMSE sebesar 22.02, MAPE sebesar 36% dan akurasi sebesar 64%. Rasio 80:20 menghasilkan RMSE sebesar 19.72, MAPE sebesar 29% dan akurasi sebesar 71%. Rasio 70:30 menghasilkan RMSE sebesar 21.87, MAPE sebesar 32% dan akurasi sebesar 68%. Rasio 60:40 menghasilkan RMSE sebesar 23.08, MAPE sebesar 35% dan akurasi sebesar 65%. Rasio 50:50 menghasilkan RMSE sebesar 23.89, MAPE sebesar 42% dan akurasi sebesar 58%. Rasio 40:60 menghasilkan RMSE sebesar 24.44, MAPE sebesar 39% dan akurasi sebesar 61%. Rasio 30:70 menghasilkan RMSE sebesar 25.26, MAPE sebesar 36% dan akurasi sebesar 64%. Rasio 20:80 menghasilkan RMSE sebesar 24.77, MAPE sebesar 43% dan akurasi sebesar 57%. Rasio 10:90 menghasilkan RMSE sebesar 27.66, MAPE sebesar 50% dan akurasi sebesar 50%.



Gambar 6. Grafik hasil pengujian random forest

Berdasarkan grafik pada gambar 6, akurasi tertinggi pada rasio 80:20 sebesar 72% dan akurasi terendah pada rasio 10:90 sebesar 50%. Rasio 80:20 memberikan proporsi yang cukup besar untuk pelatihan model (80%) sambil tetap menyisakan data yang cukup (20%) untuk menguji generalisasi model. Dengan lebih banyak data pelatihan, model dapat lebih baik mempelajari pola dan variasi dalam data. Rasio 80:20 biasanya mengurangi risiko overfitting dibandingkan rasio seperti 90:10. Jika terlalu banyak data digunakan untuk pelatihan, model mungkin belajar terlalu detail tentang data pelatihan dan tidak dapat menggeneralisasi dengan baik pada data baru.

.

4.5.3. Perbandingan Random Forest dan XGBoost

Berikut ini merupakan hasil perbandingan antara hasil pengujian menggunakan random foret dan XGBoost.

Tabel 7. Hasil perbandingan random forest dan XGBoost

Rasio Data	Random Forest	X-Boost
90:10	70%	64%
80:20	72%	72%
70:30	70%	68%
60:40	69%	65%
50:50	66%	58%
40:60	64%	61%
30:70	62%	64%

Rasio Data	Random Forest	X-Boost
20:80	58%	57%
10:90	45%	50%
Rata-Rata	64%	62%

Pada pengujian untuk peramalan random forest dan XGboost, hasil akurasi terbaik sama pada rasio data 80:20 dengan nilai akurasi yang sama sebesar 72%. Untuk rata-rata random forest memiliki rata-rata yang lebih baik sebesar 64% dibandingkan XGboost sebesar 62%. Random Forest bekerja dengan membangun banyak pohon keputusan secara paralel (bagging) dan rata-rata hasilnya. Ini memberikan stabilitas dan sering kali lebih tahan terhadap overfitting, terutama pada dataset yang lebih kecil atau dengan noise yang signifikan. XGBoost, di sisi lain, menggunakan pendekatan boosting, di mana pohon dibangun secara berurutan dengan fokus pada memperbaiki kesalahan pohon sebelumnya. Ini membuat XGBoost lebih rentan terhadap overfitting jika tidak diatur dengan baik, terutama pada dataset kecil.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi di atas, dapat disimpulkan bahwa rasio data 80:20 memberikan kontribusi yang signifikan terhadap peningkatan akurasi model, baik pada Random Forest maupun XGBoost, dengan nilai akurasi tertinggi mencapai 72% pada kedua model. Namun, jika dilihat dari rata-rata tingkat akurasi, Random Forest menunjukkan performa yang lebih stabil dengan rata-rata akurasi 64% dibandingkan XGBoost yang memiliki rata-rata akurasi sebesar 62%.

Hasil ini menunjukkan bahwa Random Forest cenderung lebih andal dalam menjaga konsistensi akurasi pada berbagai rasio data. Selain itu, perbandingan kinerja pada rasio data lainnya, seperti 10:90, mengindikasikan bahwa penurunan akurasi lebih signifikan pada Random Forest, sementara XGBoost sedikit lebih tangguh. Temuan ini memberikan wawasan bahwa pemilihan model dan rasio data yang tepat sangat penting dalam mengoptimalkan performa peramalan.

DAFTAR PUSTAKA

[1] R. N. Zacky, W. Wahyudin, and S. Dhuha, "Teknik Proyeksi Bisnis Forecasting Penjualan Menggunakan Metode Rata-Rata Trend di Storing Coffee Karawang," *J. Serambi Eng.*, vol. 8, no. 2, 2023, doi: 10.32672/jse.v8i2.5968.

[2] A. N. Rohmah and S. Subari, "Preferensi Konsumen Terhadap Produk Minuman Kopi Di Kopi Janji Jiwa Jilid 324 Surabaya," *AGRISCIENCE*, vol. 1, no. 3, 2021, doi: 10.21107/agriscience.v1i3.9129.

[3] A. C. Widiyanto, "Analisis Pengendalian Persediaan Pakan Udang Dengan Metode Min-Max Stock Pada CV. Ikhsan Jaya," *Pena J. Ilmu Pengetah. dan Teknol.*, vol. 35, no. 1, 2021, doi: 10.31941/jurnalpena.v35i1.1342.

[4] S. Audina and A. Bakhtiar, "Analisis Pengendalian Persediaan Aux Raw Material Menggunakan Metode Min-Max Stock Di PT. Mitsubishi Chemical Indonesia," *J@ti Undip J. Tek. Ind.*, vol. 16, no. 3, 2021, doi: 10.14710/jati.16.3.161-168.

[5] Z. S. Zahra and F. Fahma, "Implementasi Metode MRP untuk Pengendalian Bahan Baku Produk ABC Pada PT XYZ," *Semin. dan Konf. Nas. IDEC*, no. ISSN 2579-6429, 2020.

[6] R. Rindiyani, A. Primadewi, M. Maimunah, and A. H. Purwantini, "Klasifikasi Penjualan berdasarkan Platform pada UMKM Omah Branded Menggunakan Random Forest," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, 2022, doi: 10.30865/jurikom.v9i5.4949.

[7] M. D. A. Alhamdi, Herman, and W. Astuti, "Peramalan Kebutuhan Obat Menggunakan XGBoost Studi Kasus pada Rumah Sakit XYZ," *Indones. J. Comput. Sci.*, vol. 12, no. 5, 2023, doi: 10.33022/ijcs.v12i5.3344.

[8] S. R. Febrian *et al.*, "Prediksi Penjualan Suku Cadang Motor Dengan Penerapan Random Forest Di Pt Terus Jaya Sentosa Motor," *JATI*, vol. 8, no. 5, pp. 10507–10513, 2024.

[9] M. S. Efendi, Sarwido, and A. K. Zyen, "Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk," *RESOLUSI*, vol. 5, no. 1, pp. 12–20, 2024, doi: 10.30865/resolusi.v5i1.2149.

[10] H. Wijaya, D. P. Hostiadi, and E. Triandini, "Optimization Of Xgboost Algorithm Using Parameter Tunning In Retail Sales Prediction" *JANAPATI*, vol. 13, no. 3, pp. 769–786, 2024.

[11] Z. Aini, "Implementasi Random Forest Dan Gradient Boosting Pada Klasifikasi Indeks Pembangunan Manusia (IPM)," 2023.

[12] Y. A. Jatmiko, S. Padmadisastra, and A. Chadidjah, "Analisis Perbandingan Kinerja Cart Konvensional, Bagging Dan Random Forest Pada Klasifikasi Objek: Hasil Dari Dua Simulasi," *MEDIA Stat.*, vol. 12, no. 1, 2019, doi: 10.14710/medstat.12.1.1-12.

[13] R. Yoris, "Penggunaan Metode Xgboost Untuk Klasifikasi Status Obesitas Di Indonesia," *Univ. HASANUDDIN MAKASSAR*, 2021.

[14] E. J. Sudarman and S. Budi, "Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting untuk Prediksi Keberhasilan Studi Mahasiswa," *J. Strateg.*, vol. 5, no. 2, 2023.

[15] N. Ahmad, Y. Ghadi, M. Adnan, and M. Ali, "Load Forecasting Techniques for Power System: Research Challenges and Survey," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3187839.

[16] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14. 2022. doi: 10.5194/gmd-15-5481-2022.

[17] Ojsadmin, Adrianto Ahmad, Idral Amri, and Rahmah Nabilah, "Produksi Bioetanol Generasi Kedua dari Pelepah Kelapa Sawit dengan Variasi Pre-Treatment H2SO4 dan Waktu Fermentasi," *J. Bioprocess, Chem. Environ. Eng. Sci.*, vol. 1, no. 1, 2020, doi: 10.31258/jbchees.1.1.12-27.