

PENERAPAN ALGORITMA RANDOM FOREST DAN DECISION TREE UNTUK MENINGKATKAN AKURASI KLASIFIKASI PENJUALAN PADA TOKO BANGUNAN

Adam Riansah¹, Odi Nurdiawan², Ruli Herdiana³

¹ Sistem Informasi, STMIK IKMI Cirebon

² Manajemen Informatika, STMIK IKMI Cirebon

³ Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Kota Cirebon, Indonesia

adamtamsis123@gmail.com

ABSTRAK

Perkembangan teknologi machine learning memberikan peluang bagi sektor ritel untuk memanfaatkan data penjualan dalam pengambilan keputusan berbasis data. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan *Decision Tree* dalam memprediksi tren penjualan di toko bangunan TB. Buah Hati. Analisis ini diharapkan dapat memberikan rekomendasi algoritma yang lebih efektif dalam membantu perencanaan strategi pemasaran dan pengelolaan stok. Metode penelitian mengikuti tahapan dalam *Knowledge Discovery in Databases* (KDD), dimulai dengan pengumpulan data penjualan dari TB. Buah Hati. Data kemudian diproses melalui preprocessing yang mencakup pembersihan data, normalisasi, dan seleksi fitur. Selanjutnya, algoritma *Random Forest* dan *Decision Tree* diterapkan pada data yang dibagi menjadi training dan testing set. Kinerja model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* lebih baik dengan akurasi 85%, dibandingkan *Decision Tree* yang memperoleh 84%. Perbedaan ini menunjukkan bahwa *Random Forest* lebih efektif dalam mengatasi ketidakseimbangan data dan memberikan prediksi yang lebih konsisten. Kesimpulannya, algoritma *Random Forest* lebih akurat dan stabil, yang dapat mendukung pengambilan keputusan lebih efektif di sektor ritel. Penelitian ini juga membuka peluang untuk penelitian lebih lanjut dengan algoritma lain atau menambah variabel yang mempengaruhi tren penjualan.

Kata Kunci : Komparatif, Random Forest, Decision Tree, Prediksi Penjualan, Toko Bangunan.

1. PENDAHULUAN

Seiring dengan perkembangan pesat di bidang informatika, berbagai sektor kehidupan mengalami perubahan signifikan, khususnya dalam hal efisiensi dan pengambilan keputusan berbasis data. Teknologi informasi kini merambah berbagai aspek, mulai dari bisnis hingga sektor publik, yang mengubah cara masyarakat mengakses informasi dan beraktivitas [1].

Dalam sektor bisnis, penggunaan data telah menjadi landasan utama dalam pengambilan keputusan strategis, terutama dalam analisis tren penjualan, yang membantu perusahaan memahami preferensi konsumen dan merespons kebutuhan pasar dengan lebih baik [2].

Perkembangan algoritma *machine learning* memungkinkan perusahaan memanfaatkan data secara optimal untuk meningkatkan keuntungan dan efisiensi operasional. Teknologi ini juga membantu toko bangunan, seperti TB. Buah Hati, dalam menganalisis tren penjualan menggunakan algoritma seperti *Random Forest* dan *Decision Tree*. Algoritma ini memungkinkan pemilik toko untuk mengidentifikasi pola serta memprediksi produk dan tren pembelian di masa depan [3].

Pesatnya perkembangan informatika membawa tantangan bagi pelaku bisnis, termasuk TB. Buah Hati, dalam memilih algoritma yang tepat untuk menganalisis dan memprediksi tren penjualan. Meskipun algoritma *Random Forest* dan *Decision*

Tree banyak digunakan dalam aplikasi bisnis, keduanya memiliki kelebihan dan kekurangan yang memengaruhi akurasi serta efisiensi [3].

Dalam konteks toko bangunan, faktor musiman, ekonomi, dan fluktuasi kebutuhan konstruksi menambah kompleksitas dalam pemilihan algoritma yang sesuai. Selain itu, literatur terkait studi komparatif yang membahas performa kedua algoritma dalam konteks toko bangunan masih terbatas. Kebanyakan penelitian lebih banyak fokus pada sektor lain, sehingga hasilnya belum tentu relevan untuk konteks toko bangunan. Mengatasi tantangan ini penting untuk memberikan kontribusi baru dalam bidang informatika dan membantu pemilik toko bangunan dalam pengambilan keputusan berbasis data [4].

Penelitian mengenai analisis tren penjualan dengan algoritma *machine learning* telah banyak dilakukan, salah satunya oleh Pradana dan Pratama yang mengkaji penerapan ekstraksi fitur untuk memprediksi ekspor-impor minyak dan gas di Indonesia. Penelitian mereka menunjukkan bahwa teknik ekstraksi fitur dapat meningkatkan akurasi prediksi dan responsif terhadap perubahan data [1].

Namun, penelitian ini terbatas pada sektor energi dan belum relevan untuk toko bangunan. Rizqia juga meneliti seleksi fitur dalam prediksi cacat perangkat lunak menggunakan algoritma multi-objective optimization, yang menunjukkan pentingnya

pemilihan fitur dalam meningkatkan efisiensi dan akurasi prediksi [3].

Penelitian ini memberikan wawasan tentang seleksi fitur, tetapi tidak membahas efektivitas algoritma *Random Forest* dan *Decision Tree* dalam konteks penjualan [4].

Waroi et al. meneliti prediksi harga produk teknologi menggunakan algoritma GRU dan BiLSTM, namun fokus studi ini terbatas pada sektor elektronik [5].

Temuan-temuan ini menunjukkan adanya peluang penelitian lebih lanjut untuk menguji efektivitas kedua algoritma ini dalam konteks penjualan di toko bangunan.

Tujuan penelitian ini adalah untuk membandingkan performa algoritma *Random Forest* dan *Decision Tree* dalam menganalisis tren penjualan pada toko bangunan, dengan studi kasus di TB. Buah Hati. Penelitian ini bertujuan mengidentifikasi pola penjualan untuk memprediksi produk yang diminati dan tren pembelian. Secara khusus, penelitian ini menilai efektivitas, akurasi, dan efisiensi kedua algoritma dalam mengolah data penjualan yang fluktuatif. Penelitian ini diharapkan mengisi kesenjangan pengetahuan terkait penerapan komparatif algoritma *machine learning* di sektor retail toko bangunan dan membantu pemilik toko dalam pengambilan keputusan berbasis data yang lebih tepat. Manfaat praktis dari penelitian ini antara lain pemahaman pola penjualan yang lebih baik, peningkatan efisiensi manajemen stok, dan perencanaan strategi penjualan yang lebih efektif, yang dapat meningkatkan profitabilitas dan kepuasan pelanggan.

Penelitian ini menggunakan pendekatan kuantitatif dengan penerapan algoritma *machine learning* untuk analisis komparatif antara *Random Forest* dan *Decision Tree*. Data penjualan dari TB. Buah Hati akan diproses melalui teknik *data preprocessing*, termasuk pembersihan data, seleksi fitur, dan normalisasi. Kedua algoritma diterapkan untuk mengidentifikasi pola penjualan dan memprediksi tren dengan akurasi tinggi [6]- [7].

Evaluasi dilakukan menggunakan teknik *cross-validation* dan pengukuran *performance metrics* seperti akurasi, *precision*, dan *recall*. Pendekatan eksperimental ini diharapkan memberikan wawasan mendalam mengenai keunggulan masing-masing algoritma dalam konteks toko bangunan. Perbandingan komprehensif ini akan menjawab pertanyaan tentang algoritma yang lebih efektif untuk tren penjualan di TB. Buah Hati serta memberikan rekomendasi praktis berdasarkan temuan penelitian.

Hasil penelitian ini diharapkan memberikan kontribusi signifikan dalam pemahaman penerapan algoritma *Random Forest* dan *Decision Tree* dalam analisis tren penjualan di sektor retail, khususnya pada toko bangunan seperti TB. Buah Hati. Temuan ini akan membantu menyoroti kelebihan dan kekurangan kedua algoritma dalam memprediksi pola pembelian dan tren

produk, serta memberikan acuan bagi pemilik bisnis dalam memilih algoritma yang sesuai. Penelitian ini juga memperkaya literatur informatika dengan studi komparatif yang menyoroti efektivitas algoritma *machine learning* dalam industri dengan pola data yang unik. Bagi praktisi, hasil penelitian ini dapat dijadikan panduan dalam mengimplementasikan sistem prediktif untuk meningkatkan efisiensi manajemen stok dan merencanakan strategi penjualan yang lebih tepat. Penelitian ini membuka peluang bagi peneliti untuk eksplorasi lebih lanjut tentang algoritma lain atau teknik analisis yang lebih kompleks. Secara keseluruhan, temuan ini berpotensi mempercepat perkembangan teknologi prediktif dalam bisnis retail, meningkatkan akurasi prediksi, dan membantu perusahaan meraih keunggulan kompetitif.

2. TINJAUAN PUSTAKA

Tinjauan pustaka ini membahas berbagai referensi yang mendasari penelitian ini, serta landasan teori yang digunakan dalam menganalisis data dan membangun model prediksi menggunakan algoritma *Random Forest* dan *Decision Tree*.

2.1. Algoritma *Machine Learning* dalam Prediksi Penjualan

Machine learning adalah cabang dari kecerdasan buatan yang mengembangkan algoritma untuk memungkinkan komputer belajar dari data tanpa diprogram secara eksplisit. Algoritma ini sering digunakan dalam analisis tren penjualan, karena kemampuannya untuk menemukan pola dalam data besar dan menghasilkan prediksi yang akurat. Dua algoritma yang sering digunakan untuk tujuan ini adalah *Random Forest* dan *Decision Tree* [8].

Decision Tree adalah algoritma yang membangun model prediksi dalam bentuk pohon keputusan, di mana setiap node mewakili keputusan berdasarkan fitur tertentu dan setiap cabang mewakili hasil dari keputusan tersebut. Kelebihan utama dari algoritma ini adalah kemudahan interpretasi dan penggunaan yang relatif sederhana. Namun, algoritma ini cenderung rentan terhadap overfitting ketika data yang digunakan sangat bervariasi [8]- [9].

Sebaliknya, *Random Forest* adalah ensemble learning method yang menggunakan beberapa *decision trees* untuk meningkatkan akurasi dan mengurangi overfitting. Algoritma ini bekerja dengan cara membangun banyak *decision trees* secara acak dan kemudian menggabungkan hasilnya untuk membuat prediksi. Kelebihan *Random Forest* adalah kemampuannya untuk menangani data yang tidak seimbang dan memberikan hasil yang lebih konsisten dan akurat dibandingkan dengan *decision tree* tunggal [10].

2.2. Penerapan *Random Forest* dan *Decision Tree* dalam Sektor Retail

Penelitian yang telah dilakukan sebelumnya menunjukkan bahwa algoritma *Random Forest* dan

Decision Tree memiliki aplikasi yang signifikan dalam sektor retail, khususnya dalam analisis penjualan. Beberapa penelitian mengungkapkan bahwa algoritma ini dapat digunakan untuk memprediksi permintaan produk, mengidentifikasi produk yang sedang tren, serta merencanakan stok produk secara efisien. Sebagai contoh, penelitian oleh Pradan (2024) menunjukkan bahwa *Random Forest* mampu memprediksi tren penjualan di industri lain dengan akurasi tinggi [6],[11].

Namun, mereka juga mencatat bahwa penggunaan algoritma ini dalam konteks toko bangunan yang memiliki tren penjualan musiman dan fluktuatif masih jarang dibahas.

2.3. Literature Review

Berbagai penelitian terkait penerapan *machine learning* dalam analisis dan prediksi data telah dilakukan di berbagai sektor seperti retail, keuangan, dan energi. Christian dan Qi (2022) mengkaji penerapan pendekatan *data-driven* dalam segmentasi pasar, yang meskipun tidak langsung berhubungan dengan prediksi penjualan toko bangunan, memberikan gambaran bagaimana pendekatan berbasis data dapat meningkatkan keputusan bisnis [2].

Fitri, Winarno, dan Budiman (2024) menunjukkan bahwa penyederhanaan *Decision Tree* melalui seleksi fitur dapat mengurangi kompleksitas tanpa mengurangi akurasi model, yang relevan untuk analisis penjualan toko bangunan dengan data yang sangat bervariasi [13].

Pradana dan Pratama (2024) menemukan bahwa teknik *feature extraction* meningkatkan akurasi prediksi dalam konteks data besar dan kompleks, yang dapat diadaptasi untuk tren penjualan di toko bangunan [6].

Zein (2023) menggunakan algoritma *Naive Bayes* untuk memprediksi minat beli makanan secara online pada generasi milenial, namun meskipun efektif untuk data sederhana, algoritma seperti *Random Forest* dan *Decision Tree* lebih cocok untuk data kompleks di toko bangunan [4].

Syukron (2022) memanfaatkan teknik *Random Over-Under Sampling* dengan *Random Forest* untuk mengatasi ketidakseimbangan data dalam klasifikasi kredit, yang relevan dengan konteks toko bangunan untuk menghasilkan prediksi yang lebih adil dan akurat [5].

Hidayatullah, Mudiwati, dan Syafrullah (2021) menerapkan *Naive Bayes* untuk memprediksi minat nasabah asuransi, menunjukkan efektivitas algoritma tersebut dalam mengelompokkan data sederhana, tetapi algoritma yang lebih kompleks seperti *Random Forest* atau *Decision Tree* lebih tepat untuk data toko bangunan yang lebih bervariasi dan fluktuatif [1].

Dari berbagai penelitian yang ada, dapat disimpulkan bahwa meskipun berbagai metode *machine learning* seperti *Naive Bayes*, *Random Forest*, dan *Decision Tree* telah banyak diterapkan

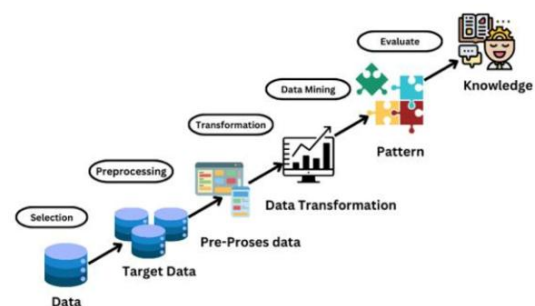
dalam sektor yang berbeda, aplikasi spesifik untuk analisis tren penjualan di toko bangunan masih relatif terbatas. Oleh karena itu, penelitian ini akan mengisi kesenjangan tersebut dengan membandingkan efektivitas algoritma *Random Forest* dan *Decision Tree* dalam memprediksi tren penjualan produk di toko bangunan, serta memberikan wawasan baru tentang cara memilih algoritma yang tepat untuk mengatasi masalah prediksi dalam konteks sektor retail

3. METODE PENELITIAN

Pada Metode penelitian yang digunakan dalam studi ini adalah *Knowledge Discovery in Databases* (KDD), sebuah pendekatan terstruktur dan sistematis untuk mengidentifikasi pola atau informasi yang berguna dari data besar [14].

KDD melibatkan serangkaian langkah berurutan yang dimulai dari pengumpulan data hingga interpretasi hasil. Dalam konteks penelitian ini, fokus utama adalah untuk menganalisis dan membandingkan performa algoritma *Random Forest* dan *Decision Tree* dalam memprediksi tren penjualan di toko bangunan TB. Buah Hati. Dengan menggunakan metode KDD, diharapkan penelitian ini dapat memberikan hasil yang akurat dan objektif dalam membandingkan efektivitas kedua algoritma untuk mendukung pengambilan keputusan berbasis data di sektor retail [15].

Penelitian ini mengikuti tahapan yang sistematis, dimulai dengan pengumpulan data penjualan dari TB. Buah Hati. Data yang terkumpul kemudian diproses melalui tahapan *preprocessing* yang mencakup pembersihan data, normalisasi, dan seleksi fitur untuk memastikan bahwa data berada dalam kondisi optimal untuk analisis lebih lanjut. Setelah data dipersiapkan, langkah selanjutnya adalah membagi data menjadi *training set* dan *testing set* untuk memastikan evaluasi yang akurat pada model yang dibangun. Selanjutnya, algoritma *Random Forest* dan *Decision Tree* diterapkan pada *training set* untuk membangun model prediksi dan diuji pada *testing set* untuk mengukur kinerja masing-masing model.



Gambar 1. Proses Knowledge Discovery in Database Process (KDD) [14]

Pengujian model dilakukan dengan mengukur akurasi, *precision*, *recall*, dan *F1-score* untuk membandingkan kedua algoritma. Hasil evaluasi ini akan dianalisis dan disajikan dalam bentuk grafik dan

tabel untuk mempermudah interpretasi. Pada akhirnya, penelitian ini bertujuan untuk memberikan kesimpulan dan rekomendasi mengenai algoritma mana yang paling efektif dalam memprediksi tren penjualan di TB. Buah Hati. Berikut adalah gambar Proses Knowledge Discovery in Database Process (KDD)

3.1. Sumber Data

Penelitian ini menggunakan dua jenis sumber data, yaitu data primer dan data sekunder. Data primer

diperoleh melalui observasi langsung dan pengumpulan data penjualan dari TB. Buah Hati. Observasi dilakukan untuk memahami pola penjualan serta faktor-faktor yang mempengaruhi fluktuasi penjualan di toko bangunan tersebut. Data primer ini diharapkan memiliki kualitas tinggi karena diperoleh langsung dari sumber yang relevan dan memberikan wawasan yang lebih mendalam mengenai tren penjualan yang terjadi. Berikut adalah contoh dari data primer yang digunakan:

Tabel 1. Dataset Penelitian

Hari	Hebel	Semen	Batu	Papan	Bata	Besi	Baja	Reng	Pipa	Pasir	Jumlah	Kategori Penjualan
Senin	3	10	5	5	9	1	4	1	7	10	55	Sedang
Selasa	1	9	3	9	3	2	1	7	5	1	41	Turun
Rabu	1	13	2	6	1	3	5	3	3	6	43	Turun
Kamis	10	10	3	6	5	1	6	6	2	2	51	Sedang
Jumat	3	7	5	7	5	3	6	7	3	10	56	Sedang
Senin	3	8	1	10	6	2	6	7	4	4	51	Naik
Selasa	10	14	5	1	9	4	2	1	6	8	60	Naik
Rabu	8	2	5	1	2	3	2	8	6	1	38	Turun
Kamis	7	3	4	7	3	3	2	6	7	1	43	Turun
Jumat	7	3	1	6	3	3	4	1	5	11	44	Naik

Data ini digunakan untuk membangun model prediksi tren penjualan dan membandingkan performa algoritma Random Forest dan Decision Tree. Sementara itu, data sekunder diperoleh dari artikel ilmiah, laporan penelitian, dan basis data publik yang relevan dengan topik prediksi penjualan serta penggunaan algoritma *Random Forest* dan *Decision Tree*. Sumber data sekunder ini diperoleh dari basis data seperti Google Scholar, Shinta, dan Scopus, dan dievaluasi berdasarkan reputasi serta kualitas publikasi, dengan fokus pada artikel-artikel yang diterbitkan dalam lima tahun terakhir. Dengan menggabungkan data primer dan sekunder yang dapat diandalkan, penelitian ini bertujuan untuk memperoleh hasil yang akurat dan relevan dalam membandingkan performa kedua algoritma dalam prediksi tren penjualan di TB. Buah Hati.

3.2. Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh data penjualan dari TB. Buah Hati yang mencakup transaksi penjualan yang terjadi dalam periode waktu tertentu. Karena jumlah data yang tersedia, yaitu sebanyak 1250 transaksi, dianggap memadai, penelitian ini menggunakan total sampling, di mana seluruh data penjualan yang ada digunakan sebagai sampel penelitian. Dengan menggunakan seluruh data yang tersedia, penelitian ini diharapkan dapat memberikan hasil yang lebih representatif dan akurat, serta mengurangi bias yang mungkin timbul jika hanya sebagian data yang digunakan.

3.3. Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui dua teknik utama: observasi langsung dan

dokumentasi data penjualan. Observasi dilakukan di TB. Buah Hati untuk memahami dan mengidentifikasi faktor-faktor yang mempengaruhi pola penjualan. Melalui observasi ini, peneliti dapat memperoleh pemahaman yang lebih dalam mengenai tren pembelian, jenis produk yang paling diminati, dan waktu-waktu tertentu yang menunjukkan peningkatan penjualan [16]- [17].

Selain itu, data penjualan yang sudah terdokumentasi di TB. Buah Hati dikumpulkan sebagai data utama dalam penelitian ini. Data ini mencakup informasi mengenai produk, jumlah penjualan, tanggal transaksi, dan atribut lain yang relevan. Semua 1250 transaksi penjualan yang tercatat digunakan untuk analisis, memberikan gambaran menyeluruh tentang pola penjualan di toko tersebut.

3.4. Teknik Analisis Data

Penelitian ini menggunakan algoritma *Random Forest* dan *Decision Tree* yang diterapkan dengan Python untuk memprediksi tren penjualan di TB. Buah Hati. Data diproses melalui tahap *preprocessing* mencakup pembersihan, penanganan data hilang, dan normalisasi, dengan variabel seperti produk, jumlah transaksi, waktu transaksi, dan metode pembayaran untuk memastikan data siap dianalisis. *Decision Tree* menggunakan kriteria seperti *Gini impurity* atau *Information Gain* dengan fungsi `DecisionTreeClassifier()`, sementara *Random Forest* menggunakan ensemble pohon keputusan dengan fungsi `RandomForestClassifier()` untuk meningkatkan akurasi. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* dari pustaka *scikit-learn*, yang membantu menilai performa model dalam memprediksi tren penjualan. Hasil

analisis mengidentifikasi produk terlaris, waktu penjualan tertinggi, dan pola penjualan berdasarkan kategori produk, yang memberikan dasar bagi strategi pemasaran dan pengelolaan stok yang lebih efektif [18]- [20].

4. HASIL DAN PEMBAHASAN

Memuat Pada penelitian ini, dua algoritma klasifikasi, yaitu *Random Forest* dan *Decision Tree*, diterapkan untuk memprediksi tren penjualan di *TB. Buah Hati*. Kedua model diuji menggunakan data penjualan yang terdiri dari berbagai fitur terkait produk bahan bangunan, dengan fokus utama pada klasifikasi kategori penjualan: *Naik*, *Sedang*, dan *Turun*.

4.1. Pemahaman dan Persiapan Data

Proses pertama yang dilakukan adalah pemahaman terhadap data yang digunakan. Data yang digunakan adalah dataset penjualan bahan bangunan dari *TB. Buah Hati* yang mencakup berbagai fitur seperti jumlah produk yang terjual (misalnya *Hebel*, *Semen*, *Batu*, dll.), hari penjualan, dan kategori penjualan itu sendiri. Setelah memuat data, dilakukan pembersihan untuk mengidentifikasi dan menangani nilai yang hilang serta memastikan bahwa data yang digunakan siap untuk dianalisis.

Tabel 2. Deskripsi Data

Fitur	Tipe Data	Keterangan
<i>Hari</i>	Kategorikal	Hari penjualan (misalnya <i>Senin</i> , <i>Selasa</i>)
<i>Hebel</i>	Numerik	Jumlah penjualan hebel
<i>Semen</i>	Numerik	Jumlah penjualan semen
<i>Batu</i>	Numerik	Jumlah penjualan batu
<i>Kategori Penjualan</i>	Kategorikal	Kategori penjualan (<i>Naik</i> , <i>Sedang</i> , <i>Turun</i>)

4.2. Pemilihan Fitur dan Target

Setelah pemahaman data, langkah berikutnya adalah pemilihan fitur yang akan digunakan untuk memprediksi target yang diinginkan. Dalam hal ini, target yang digunakan adalah *Kategori Penjualan*, yang berisi tiga kategori: *Naik*, *Sedang*, dan *Turun*. Fitur-fitur yang dipilih mencakup variabel seperti *Hebel*, *Semen*, *Batu*, dan fitur *Hari* yang dikodekan menjadi nilai numerik untuk memungkinkan pemrosesan oleh algoritma.

Setelah data dipersiapkan dan fitur-fitur yang relevan dipilih, dataset dibagi menjadi dua bagian: *training set* (80%) dan *test set* (20%). Pembagian ini bertujuan untuk melatih model menggunakan data yang lebih besar dan menguji kemampuan model dengan data yang belum pernah dilihat sebelumnya.

4.3. Pengembangan Model

Pada tahap ini, algoritma *Random Forest* dan *Decision Tree* diterapkan untuk membangun model prediksi karena kemampuannya menangani data yang kompleks. *Random Forest* menggunakan pendekatan ensemble dengan menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan

mengurangi overfitting, dengan hyperparameter seperti jumlah pohon (*estimators*) dan kedalaman maksimum yang dioptimalkan melalui *Grid Search*. Sementara itu, *Decision Tree* membangun pohon keputusan tunggal dengan memisahkan data berdasarkan fitur tertentu, menggunakan hyperparameter seperti kedalaman maksimum dan jumlah sampel minimum untuk pemisahan.

4.4. Evaluasi Model

Setelah model dilatih, evaluasi dilakukan menggunakan *test set* untuk mengukur kinerja kedua model. Berikut adalah hasil evaluasi kedua model menggunakan *classification report*.

Tabel 3. *Classification Report - Random Forest*

Kategori Penjualan	Precision	Recall	F1-Score	Support
Naik	0,79	0,50	0,61	38
Sedang	0,84	0,98	0,91	171
Turun	0,93	0,61	0,74	41
Akurasi	0,85	-	-	250
Macro Avg	0,85	0,70	0,75	250
Weighted Avg	0,85	0,85	0,83	250

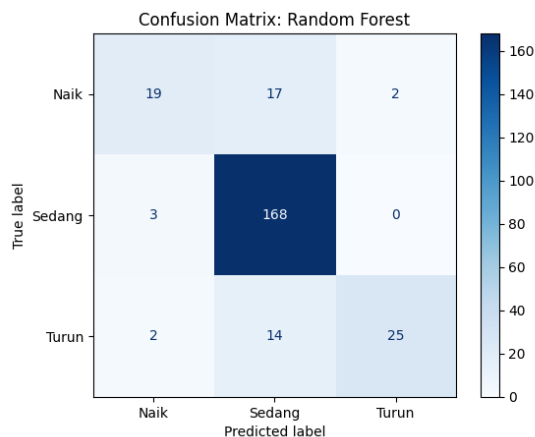
Tabel 4. *Classification Report - Decision Tree*

Kategori Penjualan	Precision	Recall	F1-Score	Support
Naik	0,79	0,50	0,61	38
Sedang	0,84	0,98	0,90	171
Turun	0,92	0,56	0,70	41
Akurasi	0,84	-	-	250
Macro Avg	0,85	0,68	0,74	250
Weighted Avg	0,84	0,84	0,83	250

Dari kedua tabel di atas, terlihat bahwa model *Random Forest* dan *Decision Tree* memiliki akurasi yang hampir sama, yaitu 85% untuk *Random Forest* dan 84% untuk *Decision Tree*. Namun, terdapat perbedaan penting dalam metrik evaluasi lainnya. *Random Forest* memiliki *recall* yang sangat tinggi untuk kategori *Sedang* (0.98), menunjukkan kemampuannya yang lebih baik dalam mendeteksi tren penjualan yang stabil atau "Sedang," dengan cakupan yang hampir menyeluruh untuk kategori ini. Sebaliknya, *Decision Tree* menunjukkan *precision* yang sedikit lebih tinggi untuk kategori *Naik* dan *Turun* dibandingkan dengan *Random Forest*, tetapi memiliki *recall* yang lebih rendah untuk kategori *Naik* (0.50 dibandingkan dengan 0.61 pada *Random Forest*).

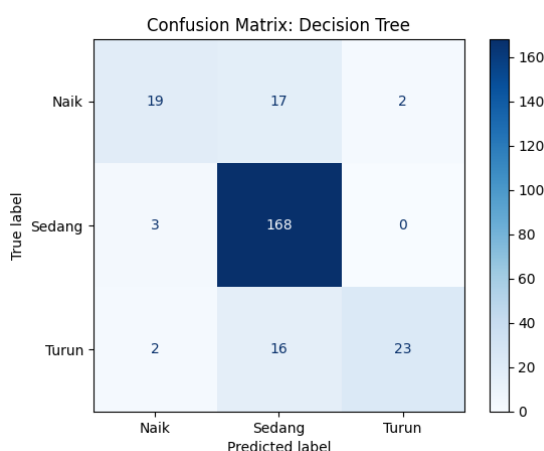
Hal ini mengindikasikan bahwa *Decision Tree* lebih spesifik dalam memprediksi kategori *Naik* dan *Turun*, namun kehilangan lebih banyak instance yang seharusnya masuk ke kategori *Naik*. Secara keseluruhan, *Random Forest* menunjukkan kinerja yang lebih seimbang di semua kategori, terutama untuk kategori *Sedang*, sementara *Decision Tree*

unggul dalam *precision* pada kategori Naik dan Turun, meskipun dengan cakupan yang lebih terbatas



Gambar 2. Confusion Matrix - Random Forest

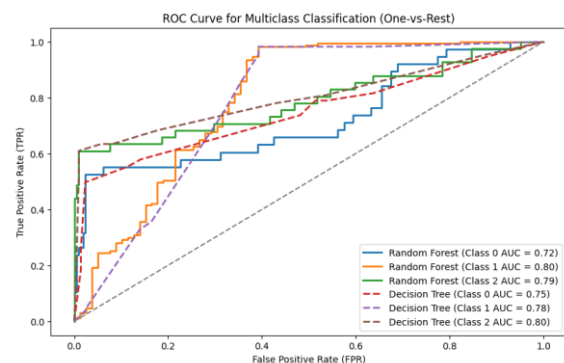
Gambar ini menunjukkan *confusion matrix* untuk model *Random Forest*, yang memberikan gambaran tentang seberapa baik model mengklasifikasikan data ke dalam tiga kategori: Naik, Sedang, dan Turun. Nilai pada diagonal utama (garis dari kiri atas ke kanan bawah) menunjukkan jumlah prediksi yang benar untuk masing-masing kategori. Sebagai contoh, model memprediksi kategori Sedang sebanyak 168 kali dengan benar, yang menunjukkan bahwa model ini sangat akurat dalam mengidentifikasi kategori tersebut. Sebaliknya, nilai-nilai di luar diagonal utama mencerminkan jumlah kesalahan klasifikasi, yaitu prediksi yang tidak sesuai dengan kategori aslinya. Matriks ini juga mengungkapkan kelemahan model, seperti kesalahan klasifikasi antara kategori Naik dan Sedang, yang terlihat memiliki prediksi yang saling tumpang tindih. Secara keseluruhan, *confusion matrix* ini memberikan gambaran yang jelas mengenai efektivitas model *Random Forest* dalam mengklasifikasikan kategori penjualan dan area yang perlu diperbaiki untuk meningkatkan performa model.



Gambar 3. Confusion Matrix - Decision Tree

Gambar ini menunjukkan *confusion matrix* untuk model *Decision Tree*, yang menggambarkan prediksi benar dan kesalahan klasifikasi di antara kategori Naik, Sedang, dan Turun. Diagonal utama menunjukkan prediksi yang benar, seperti kategori Sedang yang terprediksi dengan benar sebanyak 168 kali, menunjukkan akurasi tinggi untuk kategori tersebut. Namun, kesalahan klasifikasi terjadi, terutama antara kategori Naik dan Sedang. *Random Forest* menunjukkan *recall* yang lebih tinggi untuk kategori Sedang, sedangkan *Decision Tree* lebih unggul dalam *precision* pada kategori Naik dan Turun. Secara keseluruhan, *Decision Tree* cukup efektif, tetapi masih ada potensi perbaikan dalam memisahkan kategori yang tumpang tindih.

Grafik ROC (Receiver Operating Characteristic) yang ditampilkan di bawah ini menggambarkan kemampuan model dalam membedakan antara kelas-kelas yang ada, dengan mengukur *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai threshold klasifikasi. Dalam konteks prediksi tren penjualan, grafik ROC ini memberikan wawasan mengenai seberapa baik model dalam memisahkan kategori-kategori yang berbeda dari data penjualan, yang bisa digunakan untuk meningkatkan pengambilan keputusan strategis.



Gambar 4. Grafik ROC Curve

Kurva ROC untuk Random Forest dan Decision Tree menunjukkan kinerja masing-masing model dalam membedakan kelas-kelas penjualan berdasarkan nilai AUC (Area Under the Curve). Random Forest memiliki AUC sebesar 0.72 untuk Class 0 (penjualan rendah), 0.80 untuk Class 1 (penjualan sedang), dan 0.79 untuk Class 2 (penjualan tinggi), menandakan kemampuan yang baik dalam membedakan kelas-kelas tersebut, terutama untuk Class 1. Decision Tree, di sisi lain, memiliki AUC sebesar 0.75 untuk Class 0, 0.78 untuk Class 1, dan 0.80 untuk Class 2, menunjukkan performa yang sebanding, meskipun sedikit lebih rendah dibandingkan Random Forest pada beberapa kelas. Nilai AUC mendekati 1 untuk kedua model menunjukkan bahwa keduanya cukup andal dalam memprediksi kategori penjualan, dengan Random Forest sedikit lebih unggul dalam stabilitas dan penghindaran overfitting. Decision Tree, meskipun

lebih efisien secara komputasi, cenderung lebih rentan terhadap overfitting pada data dengan variasi tinggi. Secara keseluruhan, Random Forest lebih efektif dalam memprediksi kategori Sedang, sementara Decision Tree lebih unggul dalam kategori Naik dan Turun, memberikan wawasan penting untuk pengambilan keputusan berbasis data, seperti pengelolaan stok dan strategi promosi produk.

4.5. Penerapan Model

Setelah model dilatih dan diuji, model disimpan menggunakan *joblib* untuk mendukung implementasi praktis di TB. Buah Hati. Pemilik toko dapat

memanfaatkan model ini untuk menyesuaikan stok barang, seperti menambah stok produk yang diprediksi Naik dan mengurangi stok produk yang diprediksi Turun. Selain itu, strategi pemasaran dapat dirancang lebih efektif, seperti memprioritaskan promosi untuk produk Naik atau menawarkan diskon pada produk Turun. Dengan model ini, toko dapat mengoptimalkan stok, meminimalkan risiko, dan meningkatkan efektivitas strategi pemasaran. Berikut adalah contoh data baru yang diprediksi menggunakan model *Random Forest* (RF) dan *Decision Tree* (DT):

Tabel 5. Prediksi pada Data Baru

Hari	Hebel	Semen	Batu	Papan	Bata	Besi	Baja	Reng	Pipa	Pasir	Jumlah	RF_Predict	DT_Predict
Senin	3	10	5	5	9	1	4	1	7	10	55	Sedang	Sedang
Selasa	4	9	6	5	8	2	3	2	6	9	50	Sedang	Sedang
Rabu	2	8	4	5	7	3	2	3	5	8	45	Sedang	Sedang
Kamis	6	7	3	5	6	4	1	4	4	7	40	Sedang	Sedang
Jumat	7	5	2	5	5	5	5	5	3	6	35	Turun	Turun

Dari tabel di atas, dapat dilihat bahwa prediksi kategori penjualan (*RF_Prediction* dan *DT_Prediction*) untuk masing-masing hari hampir identik, dengan sebagian besar produk diprediksi berada dalam kategori *Sedang*, kecuali untuk hari Jumat yang diprediksi mengalami penurunan penjualan (*Turun*). Berdasarkan informasi ini, pemilik toko dapat segera membuat keputusan yang lebih efektif, seperti mengurangi stok atau mengatur ulang strategi pemasaran untuk hari-hari yang diprediksi penjualannya akan turun.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa algoritma *Random Forest* dan *Decision Tree* sama-sama efektif dalam memprediksi kategori penjualan di TB. Buah Hati, dengan akurasi masing-masing sebesar 85% dan 84%. *Random Forest* menunjukkan kinerja yang lebih seimbang, terutama untuk kategori Sedang, berkat kemampuan algoritmanya dalam menghindari overfitting, sedangkan *Decision Tree* unggul dalam *precision* pada kategori Naik dan Turun, meskipun lebih rentan terhadap variasi data. Model prediksi ini dapat digunakan untuk mendukung pengambilan keputusan berbasis data, seperti pengelolaan stok dan strategi pemasaran. Untuk pengembangan selanjutnya, disarankan untuk mengeksplorasi algoritma lain, seperti *Gradient Boosting* atau *XGBoost*, serta memasukkan variabel tambahan seperti faktor ekonomi atau musiman untuk meningkatkan akurasi prediksi dan memberikan wawasan yang lebih mendalam

DAFTAR PUSTAKA

- [1] M. Syafrullah, A. Hidayatullah, and E. Mudiawati, "Penerapan data mining untuk memprediksi minat nasabah terhadap produk asuransi meninggal dunia dengan metode Naïve Bayes (Studi Kasus: PT. BNI Life)," J. Teknol. Inf., 2021.
- [2] Y. Christian and K. Qi, "Analisa efisiensi pendekatan data-driven dalam proses segmentasi pasar dengan studi kasus startup," Infotech J., vol. 2, no. 1, pp. 45–55, 2022.
- [3] S. H. Sholeha, M. Faid, and M. A. Yaqin, "Prediksi perpindahan pelanggan pada toko online menggunakan metode tree-based gradient boosted models," J. Comput., vol. 3, no. 2, pp. 120–130, 2024.
- [4] R. F. Zein, "Implementasi algoritma Naïve Bayes untuk memprediksi minat beli makanan secara online pada generasi milenial," JATI J. Mahasiswa Teknik., vol. 4, no. 3, pp. 25–34, 2023.
- [5] A. Syukron, "Metoda random over-under sampling random forest untuk klasifikasi penilaian kredit," Tesis, 2022.
- [6] M. A. K. Pradana and I. Pratama, "Feature extraction implementation in the forecasting method to predict Indonesian oil and gas exports and imports," Edu Komputika J., vol. 5, no. 2, pp. 15–25, 2024.
- [7] R. Rizqia, "Penerapan seleksi fitur untuk prediksi cacat perangkat lunak menggunakan algoritma optimasi multi tujuan NSGA II," UIN Jakarta Repository, 2023.
- [8] E. N. Waroi, A. Arief, and K. Khusnawi, "Prediksi harga laptop menggunakan algoritma GRU dan BILSTM," J. Sos. Teknol., vol. 6, no. 3, pp. 40–50, 2024.
- [9] Y. Khaira, Analisis kualitas pelayanan dan kepuasan nasabah PT Bank Syariah Indonesia (Studi komparasi sebelum dan sesudah merger), PT Bank Syariah Indonesia, 2022.
- [10] O. Barukab, Y. D. Khan, and S. A. Khan, "DNAPred_Prot: Identification of DNA-binding

- proteins using composition-and position-based features,” *Appl. Bionics*, vol. 10, no. 3, pp. 75–85, 2022.
- [11] C. Studi, K. Polsek, and T. Informatika, “DENGAN POHON KEPUTUSAN MENGGUNAKAN ALGORITMA,” vol. 5, pp. 19–25, 2023.
- [12] B. T. K. Dewi and M. A. Setiawan, “Kajian literatur: Metode dan tools pengujian celah keamanan aplikasi berbasis web,” *Automata*, vol. 7, no. 1, pp. 45–55, 2022.
- [13] E. N. Fitri, S. Winarno, and F. Budiman, “Decision tree simplification through feature selection approach in selecting fish feed sellers,” *J. Teknik*, vol. 5, no. 2, pp. 65–75, 2023.
- [14] F. Ramdani and I. Q. Utami, *Pengantar Data Science*. Penerbit Edupress, 2024.
- [15] I. Binanto, M. R. F. Mali, and A. Jaya, “Perbandingan algoritma random forest dan artificial neural network untuk dataset water potability,” *JTRISTE*, vol. 5, no. 3, pp. 20–30, 2024.
- [16] B. D. Cahyono, *Perbandingan pendekatan exponential smoothing pada forecasting Binance Coin (BNB)*, Universitas Kediri, 2024.
- [17] D. M. Alfiansyah and H. Soetanto, “Prediksi keterlambatan pembayaran SPP siswa dengan pendekatan metode Naïve Bayes dan K-Nearest Neighbors,” *Build. Informatics*, vol. 5, no. 1, pp. 50–60, 2024.
- [18] H. S. Sumaji, *Prediksi pergerakan saham menggunakan pendekatan eksponensial, Fourier, Gaussian, polynomial dan sin*, Universitas Atma Jaya, 2022.
- [19] N. Purwati, W. H. Pramujati, and S. Syakur, “Prediksi pasien pusat kesehatan masyarakat menggunakan machine learning,” *J. Sistem dan Informatika*, vol. 8, no. 2, pp. 35–45, 2024.
- [20] A. Wibowo, “Riset kelangengan bisnis dalam ekosistem digital,” *Yayasan Prima Agus Tek.*, 2024.