ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*

SKRIPSI

DISUSUN OLEH

TRI INDAH ANGGRAINI RANGKUTI

NPM. 2109010132

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2025

ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

TRI INDAH ANGGRAINI RANGKUTI
NPM. 2109010132

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025

LEMBAR PENGESAHAN

Judul Skripsi	: ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*
Nama Mahasiswa	: TRI INDAH ANGGRAINI RANGKUTI
NPM	: 2109010132
Program Studi	: SISTEM INFORMASI

Menyetujui

Komisi Pembimbing

(Halim Maulana, S.T., M.Kom)

NIDN. 0121119102

Ketua Program Studi

(Martiano, S.Pd, S.Kom., M.Kom)

NIDN. 0128029302

Dekan

(Dr. Al-Khowarizmi, S.Kom., M.Kom)

NIDN. 0127099201

PERNYATAAN ORISINALITAS

ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, April 2025

Yang membuat pernyataan

Tri Indah Anggraini Rangkuti

NPM. 2109010132

PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Tri Indah Anggraini Rangkuti
NPM : 2109010132
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royaliti Non-Eksekutif (Non-Exclusive Royality free Right) atas penelitian skripsi saya yang berjudul:

ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royaliti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, April 2025
Yang membuat pernyataan

Tri Indah Anggraini Rangkuti
NPM. 2109010132

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Tri Indah Anggraini Rangkuti
Tempat dan Tanggal Lahir	: Padangsidimpuan, 27 Juni 2003
Alamat Rumah	: Jl. Kapt. Koima No. 135 Padangsidimpuan
Telepon/Faks/HP	: 081263962454
E-mail	: tindah945@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SD NEGERI 200107 PADANGSIDIMPUAN	TAMAT: 2015
SMP	: SMP NEGERI 3 PADANGSIDIMPUAN	TAMAT: 2018
SMA	: SMA NEGERI 1 PADANGSIDIMPUAN	TAMAT: 2021

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **"ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*"**. Skripsi ini disusun sebagai salah syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Selama penyelesaian skripsi ini, penulis banyak mendapat bimbingan, arahan, informasi, dukungan dan bantuan dari berbagai pihak. Pada kesempatan ini penulis mengucapkan terimakasih yang sebesar-besarnya kepada:

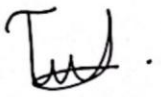
1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universias Muhammadiyah Sumatera Utara (UMSU).
2. Bapak Dr. Al- Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Bapak Martiano, S.Pd., S.Kom., M.Kom. Ketua Program Studi Sistem Informasi.
4. Ibu Yoshida Sary, S.E., S.Kom., M.Kom. Sekretaris Program Studi Sistem Informasi.
5. Bapak Halim Maulana, S.T., S.Kom, M.Kom. Dosen Pembimbing yang telah meluangkan waktu untuk memberikan saran, bimbingan serta arahan kepada penulis dalam menyelesaikan skripsi ini.
6. Kedua orang tua penulis, ayah Agustan Rangkuti dan mama Ratna Dewi Sari yang sangat penulis cintai dan sayangi. Terima kasih atas segala pengorbanan dan tulus kasih yang diberikan. Beliau memang tidak sempat merasakan pendidikan dibangku perkuliahan, namun mereka mampu senantiasa memberikan yang terbaik untuk anak-anaknya, tak kenal lelah mendoakan serta memberikan perhatian dan dukungan hingga penulis mampu menyelesaikan studinya sampai meraih gelar sarjana. Semoga ayah dan mama sehat, Panjang

umur dan bahagia selalu!.

7. Kakak dan abang penulis Eka Putri Ramadani Rangkuti, SKM dan Dwi Saputra Wijaya Rangkuti, S.H. Terima kasih atas dukungan baik secara moril maupun material yang telah diberikan kepada penulis.
8. Adik penulis, Alpin Mahmud Hasibuan dan Ardi Ansyah. Terima kasih atas keceriaan, canda, dan tawa yang selalu kalian berikan, yang menjadi penyemangat tersendiri bagi penulis selama proses penyusunan skripsi ini.
9. Sahabat terbaik penulis, Putri Qynanty pemilik NPM 2109010116. Terima kasih sudah menjadi sahabat selama perkuliahan, banyak membantu penulis dalam mengerjakan skripsi dan tidak pernah berhenti untuk saling menyemangati. *See you on top!*
10. *Last but not least,* terimakasih untuk diri sendiri Tri Indah Anggraini Rangkuti, karena telah mampu berjuang dan bertahan sejauh ini. Mampu mengendalikan diri dari berbagai tekanan diluar keadaan dan tidak memutuskan untuk menyerah sesulit apapun penulisan skripsi ini dengan menyelesaikan sebaik dan semaksimal mungkin, ini merupakan pencapaian awal yang patut dibanggakan untuk diri sendiri.

Medan, April 2025

Penulis



Tri Indah Anggraini Rangkuti

ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*

ABSTRAK

Penelitian ini fokus pada penggunaan algoritma *Random Forest* untuk memprediksi penjualan produk terlaris di PT. Arma Anugrah Abadi (Aroma Prima). Dalam menghadapi tantangan kondisi pasar yang dinamis dan persaingan yang semakin ketat, PT. Arma Anugrah Abadi (Aroma Prima) memerlukan metode yang efisien untuk memprediksi produk-produk yang memiliki potensi penjualan tertinggi. Algoritma *Random Forest* dipilih karena kemampuannya dalam memprediksi dan mengklasifikasi keputusan berdasarkan berbagai faktor seperti data historis penjualan, tren musiman, dan kondisi pasar lainnya. Data penelitian dikumpulkan melalui observasi dan wawancara di PT. Arma Anugrah Abadi (Aroma Prima), dengan implementasi menggunakan bahasa pemrograman Python pada platform Google Colab yang dapat memberikan hasil prediksi secara akurat. Berdasarkan hasil penelitian, algoritma *Random Forest* menunjukkan kinerja yang baik dalam menghasilkan keputusan.

Kata Kunci: Prediksi Penjualan, Produk Terlaris, *Random Forest*, PT.Arma Anugrah Abadi (Aroma Prima).

ANALYSIS OF BEST-SELLING PRODUCT SALES PREDICTION AT PT. ARMA ANUGRAH ABADI (AROMA PRIMA) USING RANDOM FOREST ALGORITHM

ABSTRACK

This study focuses on the use of the Random Forest algorithm to predict sales of best-selling products at PT. Arma Anugrah Abadi (Aroma Prima). In facing the challenges of dynamic market conditions and increasingly tight competition, PT. Arma Anugrah Abadi (Aroma Prima) requires an efficient method to predict products that have the highest sales potential. The Random Forest algorithm was chosen because of its ability to predict and classify decisions based on various factors such as historical sales data, seasonal trends, and other market conditions. Research data were collected through observation and interviews at PT. Arma Anugrah Abadi (Aroma Prima), with implementation using the Python programming language on the Google Colab platform which can provide accurate prediction results. Based on the results of the study, the Random Forest algorithm showed good performance in producing decisions.

Keywords: *Sales Prediction, Best-Selling Products, Random Forest, PT.Arma Anugrah Abadi (Aroma Prima).*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

PT. Arma Anugrah Abadi dengan merek dagang “Aroma Prima” adalah perusahaan yang bergerak di bidang Bakery dan Cake. Perusahaan ini memproduksi berbagai jenis roti dan kue berkualitas tinggi. Aroma Bakery dan Cake Shop memiliki jaringan toko sendiri untuk memasarkan hasil produksinya. Selain itu, Aroma Bakery juga melayani pesanan roti dari toko-toko lain yang berperan sebagai penjual roti eceran. Didirikan pada 13 Maret 2013, perusahaan ini berkomitmen untuk menjaga standar kualitas terbaik dalam setiap produknya. Saat ini, Aroma Prima Bakery & Cake Shop telah berkembang pesat dan memiliki cabang di berbagai kota, termasuk di wilayah Sumatera Utara, Nanggroe Aceh Darussalam, dan Pekanbaru.

Namun, PT. Arma Anugrah Abadi (Aroma Prima) menghadapi tantangan dalam mengelola data penjualan yang terus berkembang. Meskipun perusahaan memiliki data historis yang cukup besar, data tersebut belum dimanfaatkan secara optimal untuk mendukung pengambilan keputusan strategis. Dalam kondisi pasar yang dinamis dan persaingan yang semakin ketat, informasi tentang produk terlaris menjadi sangat penting untuk menentukan langkah-langkah efektif dalam meningkatkan penjualan dan kepuasan pelanggan.

Salah satu masalah utama PT. Arma Anugrah Abadi (Aroma Prima) adalah kesulitan dalam mengidentifikasi pola dan tren dari data penjualan yang kompleks. Hal ini sering kali menyebabkan overstock pada produk yang kurang diminati atau kekurangan stok pada produk dengan permintaan tinggi. Selain itu, strategi pemasaran yang dilakukan belum sepenuhnya didasarkan pada data yang akurat, sehingga kurang efektif dalam meningkatkan penjualan.

Dalam hal ini, algoritma *Random Forest* dapat diterapkan untuk memprediksi produk-produk yang memiliki potensi penjualan tertinggi. Algoritma ini dapat memberikan prediksi yang lebih akurat dengan mempertimbangkan berbagai faktor seperti data historis penjualan, tren musiman, dan kondisi pasar lainnya. Algoritma ini mampu mengurangi risiko kesalahan prediksi, sehingga memungkinka

pengelolaan persediaan yang lebih efisien, mengurangi biaya operasional, dan meningkatkan kepuasan pelanggan secara keseluruhan (Barros et al., n.d.).

Berdasarkan penelitian yang dilakukan (Syahrul Efendi et al., 2024) mengenai Penerapan Algoritma *Random Forest* Untuk Prediksi Penjualan Dan Sistem Persediaan Produk disimpulkan bahwa perhitungan menggunakan algoritma *Random Forest* didapatkan hasil dengan akurasi prediksi mencapai 85%. Prediksi penjualan selama seminggu ke depan menunjukkan hasil yang memadai, membantu menghindari risiko kelebihan stok yang dapat menyebabkan kerugian serta mengurangi kekurangan stok yang bisa berdampak pada penurunan kepuasan pelanggan.

Berdasarkan penelitian yang dilakukan (Febrian et al., 2024) mengenai Prediksi Penjualan Suku Cadang Motor Dengan Penerapan *Random Forest* di PT Terus Jaya Sentosa Motor disimpulkan bahwa perhitungan menggunakan algoritma *Random Forest* didapatkan hasil dengan nilai yang cukup efektif dalam memprediksi penjualan suku cadang, sehingga dapat membantu PT Terus Jaya Sentosa Motor dalam mengelola stok dengan optimal, serta mengurangi resiko kelebihan atau kekurangan persediaan.

Berdasarkan penelitian yang dilakukan (Suci Amaliah et al., 2022) mengenai Penerapan Metode *Random Forest* Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng disimpulkan bahwa perhitungan menggunakan algoritma *Random Forest* diperoleh bahwa varian minuman kategori *coffee based* lebih diminati daripada *signature coffee* dengan nilai akurasi sebesar 94,12%. Jadi, klasifikasi menggunakan metode *Random Forest* pada data penjualan minuman kopi di kedai kopi konijiwa Bantaeng dapat dikatakan akurat berdasarkan tingkat kepercayaan atau nilai akurasi yang tinggi.

Alasan peneliti menggunakan algoritma *Random Forest* karena berdasarkan uraian penelitian sebelumnya yang menggunakan algoritma tersebut untuk masalah prediksi, maka penelitian ini juga menggunakan metode *Random Forest* untuk memprediksi penjualan produk terlaris di PT. Arma Anugrah Abadi (Aroma Prima). Algoritma *Random Forest* merupakan sebuah algoritma untuk melakukan prediksi dan klasifikasi (Apriliah et al., 2021).

Dengan adanya prediksi penjualan produk terlaris menggunakan algoritma *Random Forest* maka dapat mempermudah penjualan di PT. Arma Anugrah Abadi (Aroma Prima). Berdasarkan latar belakang di atas penulis tertarik untuk melakukan penelitian dengan judul **"ANALISIS PREDIKSI PENJUALAN PRODUK TERLARIS DI PT. ARMA ANUGRAH ABADI (AROMA PRIMA) MENGGUNAKAN ALGORITMA *RANDOM FOREST*"**.

1.2. Rumusan Masalah

Permasalahan yang akan diteliti dan dibahas dalam Tugas Akhir ini yaitu bagaimana memprediksi penjualan produk terlaris di PT. Arma Anugrah Abadi (Aroma Prima) menggunakan algoritma *Random Forest*?

1.3. Batasan Masalah

Berdasarkan permasalah yang ada, maka penulis membatasi masalah pada penelitian ini antara lain:

1. Algoritma yang digunakan untuk menentukan prediksi adalah *Random Forest.*
2. Data yang diambil adalah data penjualan PT. Arma Anugrah Abadi (Aroma Prima) tahun 2024.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini agar mempermudah PT. Arma Anugrah Abadi (Aroma Prima) memprediksi penjualan produk terlaris yang terjual dan juga sebagai sumber informasi penting untuk meningkatkan penjualan di PT. Arma Anugrah Abadi (Aroma Prima).

1.5. Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Bagi Peneliti

   Meningkatkan pengetahuan terkait penerapan algoritma *Random Forest* dalam memprediksi penjualan produk terlaris di PT. Arma Anugrah Abadi (Aroma Prima).

2. Bagi Perusahaan

   Memudahkan perusahaan dalam memprediksi produk terlaris secara cepat dan tepat.

3. Bagi Universitas

   Memperoleh referensi tingkat baru dan suatu karya tulis untuk memenuhi suatu kewajiban mahasiswa strata 1 (satu) dalam menyelesaikan perkuliahan.

BAB II
LANDASAN TEORI

2.1. Prediksi

Menurut Kamus Besar Bahasa Indonesia (KBBI), prediksi adalah hasil dari aktivitas memprediksi, meramal, atau memperkirakan sesuatu. Ini sama dengan ramalan atau perkiraan mengenai nilai yang akan datang dengan menggunakan data yang diperoleh dari masa lalu. Prediksi berfungsi untuk menunjukkan kemungkinan yang akan terjadi dalam suatu keadaan tertentu dan menjadi input dalam proses perencanaan serta pengambilan keputusan. Prediksi dapat didasarkan pada metode ilmiah atau sekedar opini subjektif.

Prediksi sendiri adalah perkiraan tentang kejadian yang mungkin terjadi di masa depan. Dalam bidang ilmu sosial, ketidakpastian seringkali membuat prediksi menjadi sulit dilakukan secara akurat. Tujuan dari prediksi adalah untuk mengurangi dampak ketidakpastian dan meminimalkan kesalahan dalam peramalan (Najla et al., n.d.).

Prediksi berfokus pada upaya memperkirakan nilai suatu variabel di masa depan. Terdapat tiga jenis prediksi, yakni jangka pendek, menengah, dan panjang. Prediksi jangka pendek berfokus pada pola data dalam periode yang singkat, sementara prediksi jangka menengah dan panjang lebih digunakan untuk perencanaan strategis, seperti ekspansi atau perencanaan kebutuhan jangka panjang (Dzickrillah Laksmana et al., 2019).

2.2. Penjualan

Penjualan adalah aktivitas ekonomi yang terjadi ketika perusahaan memperoleh hasil atau manfaat yang telah direncanakan, baik dalam bentuk keuntungan dari penjualan atau pengembalian biaya yang telah dikeluarkan (Eka Pratiwi et al., 2022).

Penjualan produk perusahaan dihitung setelah dikurangi dengan potongan harga dan pengembalian barang (retur). Secara umum, penjualan merujuk pada transaksi di mana satu pihak membeli barang atau jasa dari pihak lain dengan pembayaran berupa uang. Berdasarkan pengertian tersebut, dapat disimpulkan bahwa penjualan adalah aktivitas ekonomi yang melibatkan pembeli dan penjual,

di mana terjadi pertukaran barang atau jasa dengan pembayaran uang tunai (Andi Saputra et al., 2020).

2.3. *Machine Learning*

Machine learning merupakan salah satu cabang dari kecerdasan buatan yang memungkinkan sistem meniru kemampuan manusia dalam belajar dari pengalaman. Dalam penerapannya, *machine learning* memanfaatkan algoritma atau serangkaian proses statistik untuk menghasilkan prediksi berdasarkan pengolahan data menggunakan prinsip-prinsip statistik. Data tersebut kemudian digunakan sebagai dasar dalam pengambilan keputusan dan pembuatan prediksi. Kualitas algoritma yang digunakan sangat memengaruhi tingkat akurasi hasil prediksi dan keputusan sistem (Mahendra et al., 2022). Beberapa manfaat dari penerapan *machine learning* antara lain:

a. Klasifikasi *(Classification)*, teknik ini digunakan untuk memprediksi nilai atau kategori individu dalam suatu populasi.
b. Pencocokan Kemiripan *(Similarity Matching)*, merupakan teknik yang digunakan untuk mengidentifikasi tingkat kemiripan antar individu berdasarkan data yang tersedia.
c. Pengklasteran *(Clustering)*, teknik ini berfungsi untuk mengelompokkan individu ke dalam kelompok tertentu berdasarkan kesamaan karakteristik yang dimiliki.

2.4 *Ensamble Learning*

Menurut (Nguyen et al., 2021) *Ensemble Machine Learning* merupakan pendekatan yang mengombinasikan beberapa model pembelajaran mesin, baik yang sejenis (homogen) maupun berbeda jenis (heterogen), dengan tujuan meningkatkan akurasi prediksi serta mengurangi pengaruh noise atau kesalahan dari data yang dianalisis dan diprediksi. Umumnya, metode *ensemble* dibagi ke dalam tiga jenis utama, yaitu b*agging (bootstrap aggregating)*, *boosting,* dan *squared.* Ketiga pendekatan ini menyesuaikan hasil prediksi berdasarkan perhitungan gabungan dari semua model, baik dalam hal bias, variansi, atau keduanya secara bersamaan. Perbedaan utama di antaranya terletak pada jenis model yang digunakan: *bagging* dan *boosting* biasanya memakai model homogen,

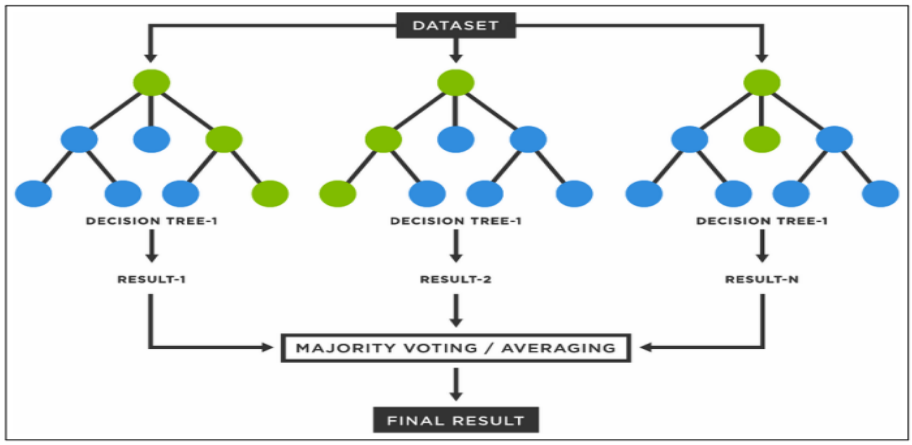
sedangkan *squared* lebih optimal dalam menggabungkan model-model yang heterogen.

2.5 Algoritma *Random Forest*

Random Forest adalah algoritma yang diperkenalkan oleh Leo Breiman pada tahun 2001 dan termasuk dalam metode *machine learning* yang sangat efektif untuk menyelesaikan permasalahan klasifikasi maupun regresi (Apriliah et al., 2021). Algoritma ini terdiri dari kumpulan pohon keputusan *(decision trees)* yang bekerja secara bersamaan untuk menghasilkan keputusan, serta memiliki kinerja yang baik, terutama pada dataset berukuran besar. *Random Forest* juga mampu menangani data dengan nilai yang hilang tanpa perlu menghapus variabel tersebut (Ullah et al., 2019).

Perbedaan mendasar antara *Random Forest* dan *Decision Tree* terletak pada cara menghasilkan prediksi. Jika *Decision Tree* hanya mengandalkan satu pohon keputusan yang cenderung rentan terhadap *overfitting*, maka *Random Forest* menggunakan banyak pohon secara bersamaan, dan menentukan hasil akhir melalui proses agregasi seperti voting mayoritas untuk klasifikasi atau rata-rata untuk regresi. Pendekatan ini, yang memanfaatkan teknik *bagging* dan pemilihan fitur secara acak, menjadikan *Random Forest* lebih andal dalam menghindari *overfitting* dan menngkatkan kemampuan generalisasi model.

Secara operasional, *Random Forest* membentuk banyak pohon keputusan, kemudian menggabungkan hasil dari masing-masing pohon untuk menghasilkan prediksi akhir, sehingga mampu mengatasi kelemahan prediksi yang sering terjadi bila hanya mengandalkan satu pohon (Sari et al., 2023).



Gambar 2.1 *Random Forest*

Gambar 2.1 menggambarkan alur proses kerja dari algoritma *Random Forest*, yang terdiri dari tahapan-tahapan sebagai berikut:

1. Algoritma terlebih dahulu memilih data secara acak sebagai sampel.
2. Untuk setiap sampel acak yang diambil, algoritma membangun pohon keputusan (*decision tree*) guna menghasilkan nilai prediksi dari masing-masing pohon.
3. Setelah seluruh pohon terbentuk dan menghasilkan prediksi, dilakukan penggabungan hasil: untuk masalah klasifikasi digunakan nilai yang paling sering muncul (modus), sedangkan untuk regresi digunakan nilai rata-rata (mean).
4. Algoritma kemudian memilih hasil akhir berdasarkan pendekatan yang sesuai untuk memperoleh prediksi terbaik.

Proses pembentukan *Random Forest* diawali dengan pengambilan sejumlah dataset acak sebanyak *n* (jumlah total data) dari data latih, menggunakan teknik *bootstrap sampling*, yaitu pengambilan sampel dengan penggantian (*replacement*), sehingga satu data dapat terpilih lebih dari sekali atau bahkan tidak terpilih sama sekali. Kemudian, algoritma CART (*Classification and Regression Trees*) diterapkan untuk menyusun pohon keputusan dari setiap dataset acak tersebut dengan menggunakan aturan pemisahan *(splitting)*, seperti indeks Gini, Entropy, atau kriteria lain yang sesuai. Setiap pohon yang terbentuk dari proses ini menjadi bagian dari keseluruhan struktur *Random Forest.*

$$Gini = 1 - \sum_{i=1}^{n} p_i^2$$

Dimana:

Gini(S) : Nilai Gini untuk dataset *S*.

n : Jumlah kelas dalam dataset.

Pi : Probabilitas suatu kelas dalam dataset *S*.

$$Entropy = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

Dimana:

$Entropy(S)$: Entropy dari dataset *S*.

n : Jumlah kelas dalam dataset.

Pi : Probabilitas suatu kelas dalam dataset *S*.

2.6 Penelitian Terdahulu

Dalam menyusun penelitian ini dapat dikaitkan dengan penelitian sebelumnya. Berikut adalah beberapa penelitian yang relevan dengan penelitian ini:

Tabel 2.1. Penelitian Terdahulu

No.	Judul	Penulis	Tahun	Kesimpulan
1.	Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk	(Syahrul Efendi et al.)	2024	Penerapan Algoritma *Random Forest* dalam melakukan prediksi penjualan dan sistem persediaan produk Bolen Crispy berhasil dilakukan, dengan akurasi prediksi mencapai 85%. Ini menunjukkan bahwa model ini dapat diandalkan dalam memproyeksikan penjualan dengan ketepatan yang cukup baik.
2.	Prediksi Penjualan Suku Cadang Motor Dengan Penerapan Random Forest Di PT Terus Jaya Sentosa Motor	(Febrian et al.)	2024	Penerapan Algoritma *Random Forest* dalam prediksi penjualan suku cadang motor di PT Terus Jaya Sentosa Motor berhasil dilakukan. Menunjukkan hasil evaluasi dengan RMSE sebesar 23.37722, MAE

				sebesar 13.96572 serta nilai MSE sebesar 546.49478, hasil ini menandakan nilai RMSE menunjukkan rata-rata perbedaan prediksi dari nilai sebenarnya sekitar 23.38 unit, sedangkan rata-rata prediksi yang meleset sekitar 13.97 unit dari nilai sebenarnya. Meskipun dengan nilai MSE yang besar model ini dapat dikatakan cukup baik karena nilai RMSE dan MAE yang kecil.
3.	Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng	(Suci Amaliah et al.)	2022	Penerapan Algoritma Random Forest dalam melakukan klasifikasi varian minuman kopi di kedai kopi konijiwa Bantaeng berhasil dilakukan. Berdasarkan hasil analisis diperoleh bahwa model dengan error klasifikasi terkecil adalah dengan menggunakan mtry 2 dan ntree 500. Model yang dihasilkan dievaluasi dengan menggunakan confusion matrix dimana diperoleh bahwa varian

				minuman kategori *coffee based* lebih diminati daripada *signature coffee* dengan nilai akurasi sebesar 94,12%. Jadi, klasifikasi menggunakan metode *Random Forest* pada data penjualan minuman kopi di kedai kopi konijiwa Bantaeng dapat dikatakan akurat berdasarkan tingkat kepercayaan atau nilai akurasi yang tinggi.

2.7 Python

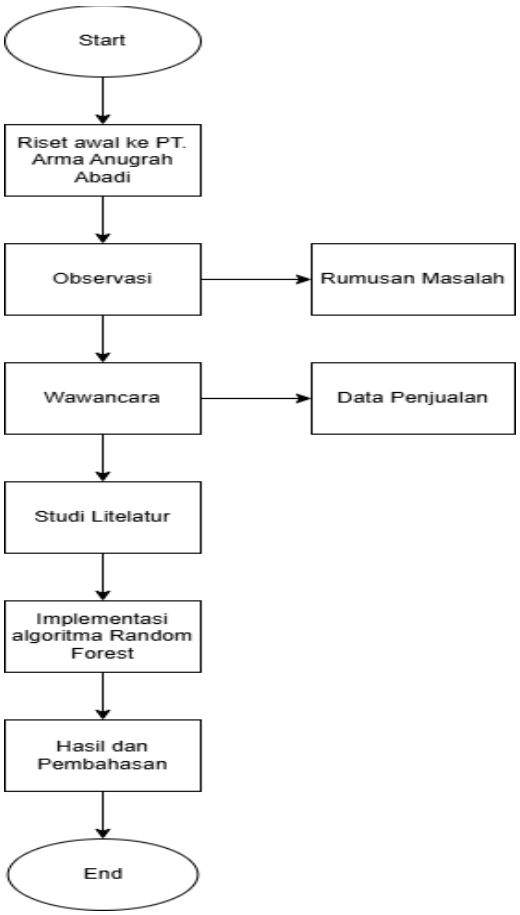
Python merupakan bahasa pemrograman tingkat tinggi yang bersifat *interpreted*, serba guna, dan memiliki sintaks yang sederhana serta mudah dipahami (Python, 2023). Dikembangkan pertama kali oleh Guido van Rossum pada tahun 1991, bahasa ini terus berkembang pesat dan kini menjadi salah satu bahasa pemrograman yang paling banyak digunakan di dunia. Python terkenal dengan desainnya yang bersih dan mudah dibaca, serta mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek (OOP), fungsional, dan prosedural. Berkat komunitas yang besar dan ekosistem pustaka yang luas, Python banyak digunakan dalam berbagai bidang, seperti pengembangan web, analisis data, kecerdasan buatan (AI), serta pemrograman sistem (Alfarizi et al., 2023).

BAB III
METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah kuantitatif. Penelitian kuantitatif merupakan sebuah mekanisme menemukan pengetahuan dan mempergunakan data yang berbentuk angka sebagai alat dalam menganalisis eksplanasi apa yang ingin diketahui. Angka menjadi landasan utama selain kriteria yang paling diperlukan pada penelitian ini.

Pendekatan kuantitatif juga merupakan pendekatan yang berfokus pada pengumpulan dan analisis data numerik. Pendekatan ini cocok digunakan dalam penelitian yang melibatkan algoritma *Random Forest*, karena algoritma tersebut memerlukan data kuantitatif untuk melakukan perhitungan dan analisis secara efektif.



Gambar 3.1 Tahapan Penelitian

Berikut penjelasan mengenai tahapan penelitian berdasarkan diagram alir yang kamu berikan:

1. Start

   Tahap awal dimulainya proses penelitian.

2. Riset Awal ke PT. Arma Anugrah Abadi

   Peneliti melakukan pengenalan awal terhadap objek penelitian, yaitu PT. Arma Anugrah Abadi (Aroma Prima). Tujuannya adalah memahami proses bisnis, sistem penjualan, dan permasalahan yang sedang dihadapi oleh perusahaan.

3. Observasi

   Peneliti mengamati langsung aktivitas penjualan di perusahaan untuk mengetahui alur transaksi, jenis data yang tersedia, dan permasalahan nyata di lapangan. Dan untuk menemukan potensi permasalahan dan merumuskan masalah penelitian.

4. Wawancara

   Dilakukan wawancara kepada pihak terkait (seperti bagian operasional atau manajemen) untuk menggali informasi lebih dalam, terutama mengenai data penjualan dan kendala yang dihadapi dalam pengelolaan stok atau analisis produk terlaris.

5. Studi Literatur

   Peneliti mengkaji referensi ilmiah berupa jurnal, buku, dan skripsi sebelumnya untuk memperkuat teori dan metode yang akan digunakan. Ini juga mendukung pemilihan algoritma *Random Forest* sebagai solusi.

6. Implementasi Algoritma Random Forest

   Tahap teknis di mana data penjualan yang telah diperoleh diolah menggunakan algoritma *Random Forest* dengan bantuan tools seperti Python dan *Scikit-learn* untuk menghasilkan prediksi produk terlaris.

7. Hasil dan Pembahasan

   Hasil dari implementasi algoritma dianalisis menggunakan metrik evaluasi (MAE, RMSE, R^2 Score), lalu dilakukan pembahasan terkait faktor-faktor yang mempengaruhi penjualan, serta bagaimana hasil ini dapat diaplikasikan dalam strategi bisnis.

8. End

   Penelitian selesai setelah seluruh tahapan dilakukan dan kesimpulan serta saran telah dirumuskan.

3.2. Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di tempat yang relevan dan dalam rentang waktu yang telah ditentukan, sesuai dengan tujuan dan kebutuhan penelitian. Penjelasan mengenai tempat dan waktu penelitian dapat dilihat pada subbab berikut:

3.2.1 Tempat Penelitian

Tempat penelitian untuk penerapan algoritma *Random Forest* dalam pengelompokkan prediksi produk terlaris dilaksanakan di PT. Arma Anugrah Abadi (Aroma Prima) tepatnya di Jl. Panglima Denai, Amplas, Kec. Medan Amplas, Kota Medan, Sumatera Utara 20229.

3.2.2 Waktu Penelitian

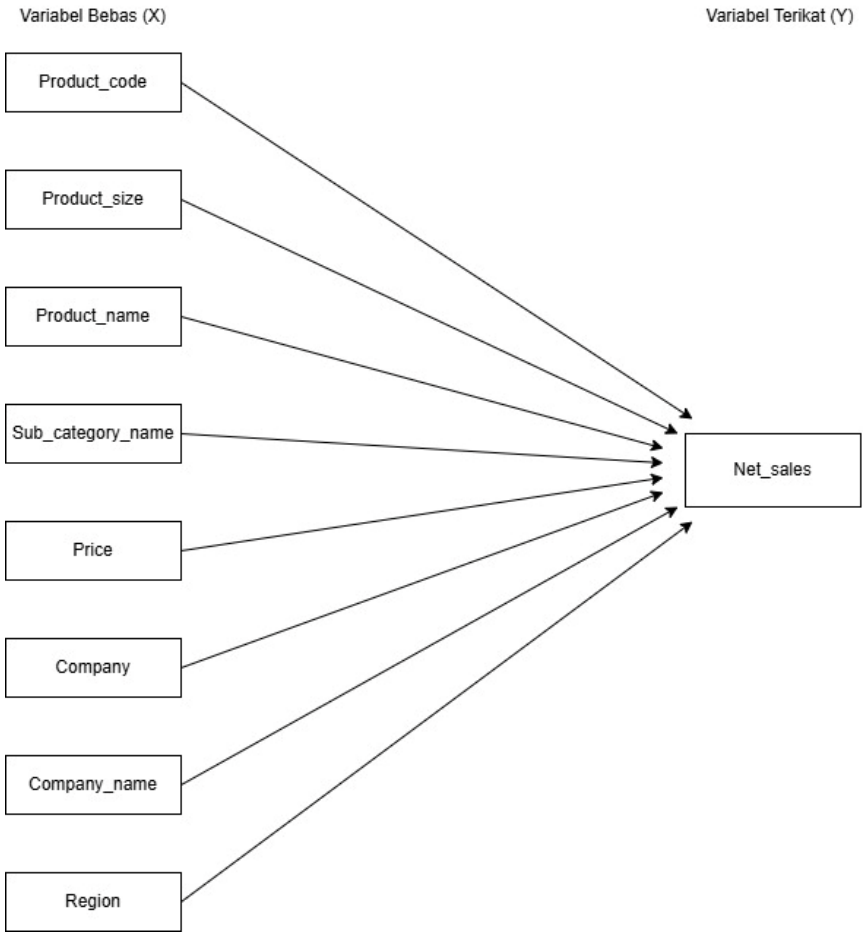
Waktu penulis dihabiskan untuk penelitian ini, dimulai pada Desember 2024 dan berakhir pada April 2025. Adapun schedule penelitian yang dilaksanakan yaitu sebagai berikut:

Tabel 3.1. *Schedule* Penelitian

No.	Aktivitas Penelitian	Waktu				
		Desember	Januari	Februari	Maret	April
1.	Pemberian Judul					
2.	Mengumpulkan Informasi					
3.	Perapian					
4.	Arahan Proposal					
5.	Seminar Proposal					
6.	Perapihan Skripsi					
7.	Arahan Skripsi					
8.	Sidang Meja Hijau					

3.3 Kerangka Konseptual

Kerangka konsep dalam penelitian ini bertujuan untuk menjelaskan hubungan antara variabel-variabel yang digunakan dalam memprediksi penjualan produk terlaris menggunakan *Random Forest*.



Gambar 3.2 Kerangka Konseptual

Kerangka konseptual pada gambar tersebut menggambarkan hubungan antara variabel bebas (X) dan variabel terikat (Y) dalam penelitian prediksi penjualan produk terlaris di PT. Arma Anugrah Abadi (Aroma Prima) menggunakan algoritma *Random Forest.*

a. Variabel Bebas (X)

Variabel bebas merupakan atribut atau fitur yang digunakan sebagai input dalam model prediksi. Dalam konteks ini, variabel bebas terdiri dari:

1. Product_code

   Merupakan kode unik yang merepresentasikan masing-masing produk.

2. Product_size

   Menunjukkan ukuran dari produk, misalnya kecil, sedang, atau besar.

3. Product_name

   Nama dari produk yang dijual.

4. Sub_category_name

   Kategori atau sub-kategori produk, seperti roti manis, kue ulang tahun, dll.

5. Price

   Harga produk per unit. Variabel ini sangat penting karena berpengaruh langsung terhadap volume penjualan.

6. Company

   ID atau kode perusahaan atau cabang penjual produk.

7. Company_name

   Nama perusahaan atau unit cabang tempat produk dijual.

8. Region

   Lokasi atau wilayah geografis penjualan yang mungkin memengaruhi minat konsumen.

b. Variabel Terikat (Y)

Variabel terikat adalah hasil yang ingin diprediksi, yaitu:

1. Net_sales

   Merupakan target atau output dari model yang ingin diprediksi, yaitu jumlah penjualan bersih dari suatu produk dalam satuan kuantitas atau nominal. Variabel ini akan dipengaruhi oleh semua variabel bebas di atas.

3.4 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan untuk memastikan bahwa proses analisis berjalan dengan baik dan menghasilkan output yang valid. Adapun tahapan-tahapan tersebut adalah sebagai berikut:

3.4.1 Teknik Pengumpulan Data

Teknik pengumpulan data merupakan tahapan yang sangat penting dalam sebuah penelitian. Teknik pengumpulan data yang benar akan menghasilkan data yang memiliki kredibilitas tinggi, begitu pula sebaliknya. Oleh karena itu, tahapan ini tidak bisa salah dan harus dilakukan secara hati-hati sesuai dengan prosedur dan karakteristik penelitian kuantitatif. Sebab, kesalahan atau ketidaksempurnaan

dalam metode pendataan akan berakibat fatal, berupa data yang tidak kredibel, sehingga hasil penelitian tidak dapat dipertanggungjawabkan. Teknik pengumpulan data adalah cara yang digunakan oleh peneliti untuk mengumpulkan data penelitian dari sumber data (subjek dan sampel penelitian).

Dalam penelitian ini penulis memanfaatkan tiga teknik pengumpulan data, yaitu studi observasi, wawancara, dan studi literatur. Peneliti menjelaskan sebagai berikut:

1. Observasi

Melalui metode ini penulis terjun ke lapangan untuk meminta izin kepada pihak PT. Arma Anugrah Abadi (Aroma Prima) untuk meneliti di perusahaan tersebut. Dan melakukan pengamatan di PT. Arma Anugrah Abadi (Aroma Prima) setelah mendapatkan izin untuk melaksanakan penelitian.

2. Wawancara

Wawancara dilakukan pada bagian Operasional PT. Arma Anugrah Abadi (Aroma Prima) untuk mengumpulkan data penelitian. Berdasarkan hasil wawancara, penulis dapat merumuskan beberapa variabel atau atribut yang akan digunakan dalam penelitian data penjualan produk.

3. Studi Literatur

Dalam studi litelatur, penulis memperoleh bahan penulisan dengan membaca buku tentang penelitian ilmiah, jurnal yang relevan dengan kasus yang sedang diteliti, serta sumber lain yang terkait dengan algoritma *Random Forest*. Serta beberapa penelitian yang berkaitan dengan prediksi penjualan produk terlaris.

3.4.2 Pemeriksaan Data *(Assessing Data)*

Tahap ini dilakukan untuk memeriksa dan identifikasi masalah yang terdapat dalam data dan memastikan data tersebut memiliki kualitas yang baik. Pemeriksaan data akan dilakukan dengan bantuan library pandas. Menurut Dicoding Team (2023), adapun beberapa masalah umum yang biasanya dijumpai dalam sebuah data adalah sebagai berikut:

1. *Missing Value*

Masalah ini muncul ketika adanya nilai yang hilang dari sebuah data dan biasanya direpresentasikan dengan nilai NaN dalam library pandas. Library pandas sendiri menyediakan metode Bernama isnull() atau isna() untuk mengidentifikasi

masing masing nilai yang hilang dalam sebuah data yang kemudian dipadukan dengan metode sum() untuk menghitung jumlah missing value dalam data.

2. *Invalid Value*

Masalah ini muncul ketika terdapat nilai yang tidak masuk akal dan tidak sesuai dengan ketentuan.

3. *Duplicate Value*

Masalah ini terjadi Ketika terdapat data yang memiliki nilai yang sama persis pada setiap kolomnya. Library pandas menyediakan metode duplicated() untuk mengidentifikasi apakah terdapat duplikasi terhadap data

4. *Inaccurate Value*

Masalah ini muncul ketika nilai pada sebuah data tidak sesuai dengan hasil observasi. Masalah ini umumnya muncul karena adanya human error dalam pencatatan transaksi.

5. *Inconsistent Value*

Masalah ini muncul Ketika sebuah data memiliki nilai yang tidak konsisten baik dari segi satuan maupun ketentuan penilaian.

3.4.3 Pembersihan dan Transformasi Data *(Cleaning Data)*

Pembersihan data dilakukan berdasarkan temuan-temuan masalah yang ditemui pada proses pemeriksaan data, masalah-masalah yang ditemukan ditangani dengan teknik-teknik tertentu yang akan dibantu menggunakan *library pandas*. Sedangkan tahapan transformasi data merupakan tahapan dimana data yang sudah melewati tahap pemeriksaan dan pembersihan diubah dengan membuat ringkasan (agregasi) sehingga data tersebut dapat digunakan dalam proses data mining menggunakan metode *Random Forest.*

Berikut beberapa tahapan penerapan pengolahan data tersebut menggunakan library pandas:

1. Mengimpor Library Pandas

   Langkah ini pertama kali dilakukan agar dapat menggunakan fitur-fitur yang disediakan pandas untuk pemrosesan data, adapun potongan kode yang digunakan dalam mengimpor Library Pandas dalam lingkungan pengembangan Python dapat dilihat pada gambar dibawah ini:

```
import padas as pd
```

Gambar 3.3 Kode Import Library Pandas

Potongan kode di atas adalah contoh penggunaan pernyataan import untuk mengimpor library Pandas dengan alias pd. Alias digunakan untuk mempermudah pemanggilan saat pandas digunakan dalam pemrosesan data.

2. Memasukkan data dalam format csv

Pandas memberikan kemudahan untuk memasukkan data kedalam lingkungan pengembangan python, adapun memasukkan data ke dalam lingkungan pengembangan model menggunakan bahasa python dan library pandas dapat dilihat dalam potongan kode dibawah ini:

```
# Memuat file CSV ke dalam DataFrame
data = pd.read_csv('nama_file.csv')
```

Gambar 3.4 Kode memuat file CSV ke dalam DataFrame

Setelah kode ini dieksekusi, data dari file CSV akan dimuat ke dalam objek DataFrame yang disebut data, yang dapat anda gunakan untuk melakukan berbagai operasi pemrosesan data menggunakan Pandas.

3. Verifikasi data

Pemeriksaan apakah ada data yang hilang atau data tidak valid dapat dipermudah dengan menggunakan library pandas. Pengecekan nilai yang hilang dalam suatu data dapat dilakukan dengan potongan kode yang terdapat di gambar3.5 dibawah ini.

```
# Misalnya, identifikasi nilai yang hilang
data.dropna(inplace=True)
```

Gambar 3.5 Kode identifikasi nilai yang hilang

Potongan kode di atas menunjukkan cara menangani data yang memiliki nilai hilang *(missing values)* menggunakan library pandas dalam bahasa pemrograman Python.

3.5 Penerapan Algoritma Random Forest

1. Gini

   Gini digunakan untuk mengukur impurity (ketidakmurnian) suatu dataset. Rumusnya:

$$Gini = 1 - \sum_{i=1}^{n} p_i^2$$

   Dimana:

   Gini(S) : Nilai Gini untuk dataset *S*.

   n : Jumlah kelas dalam dataset.

   Pi : Probabilitas suatu kelas dalam dataset *S*.

2. Entropy

   Entropy digunakan untuk menentukan homogenitas dari sebuah sampel data. Rumusnya adalah:

$$Entropy = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

   Dimana:

   $Entropy(S)$: Entropy dari dataset *S*.

   n : Jumlah kelas dalam dataset.

   Pi : Probabilitas suatu kelas dalam dataset *S*.

   Dengan menggunakan 2 rumus di atas, maka dapat dilakukan tahapan dari pengolahan data.

3.6 Evaluasi dan Validasi

Tahapan ini digunakan untuk mengukur performa dari algoritma *Random Forest* yang diterapkan pada data. Adapun metode evaluasi yang dilakukan adalah dengan metode *Confusion Matrix* dengan menghitung akurasi. Akurasi adalah proporsi prediksi benar (baik positif maupun negative) terhadap total jumlah data yang diuji. Dihitung dengan rumus:

Gambar 3.6 Confusion Matrix

Dari gambar di atas, maka diperoleh turunan rumus sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

Dimana:

TP (True Positive) : Jumlah prediksi benar untuk kategori positif.

TN (True Negative) : Jumlah prediksi salah untuk kategori positif.

FP (False Positive) : Jumlah prediksi benar untuk kategori negatif.

FN (False Negative) : Jumlah prediksi salah untuk kategori negatif.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Pendahuluan

Pendahuluan ini memberikan gambaran umum mengenai penerapan algoritma *Random Forest* dalam penelitian ini. Fokus utama adalah pada latar belakang pengimplementasian algoritma dan alur proses yang dilalui dalam penerapannya untuk analisis prediksi produk terlaris.

4.1.1 Latar Belakang Implementasi

Random forest adalah algoritma yang dikembangkan oleh Leo Breiman (2001) dan merupakan salah satu metode dalam *machine learning* yang sangat efektif untuk menangani masalah klasifikasi dan regresi. Algoritma ini terdiri dari kumpulan decision tree, yang bekerja secara bersamaan untuk meningkatkan akurasi prediksi. Salah satu keunggulan utama dari *Random Forest* adalah kemampuannya dalam menangani dataset besar serta mengatasi nilai yang hilang tanpa harus menghapus variabel terkait.

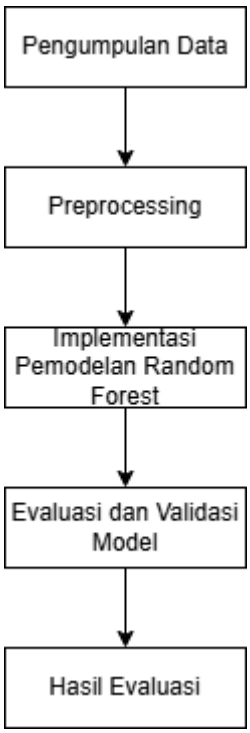
Keunggulan lain dari *Random Forest* adalah kemampuannya dalam mengurangi risiko kesalahan prediksi, sehingga dapat membantu bisnis dalam:

1. Mengoptimalkan manajemen persediaan, dengan memproduksi atau menyimpan lebih banyak produk yang diprediksi laris.
2. Mengurangi biaya operasional, dengan menghindari kelebihan stok atau kekurangan stok.
3. Meningkatkan kepuasan pelanggan, dengan memastikan ketersediaan produk sesuai permintaan pasar.

Dengan berbagai keunggulan tersebut, *Random Forest* menjadi metode yang tepat untuk diterapkan dalam analisis prediksi produk terlaris, karena memberikan keseimbangan antara akurasi, efisiensi, dan fleksibilitas dalam menangani berbagai jenis data.

4.1.2 Alur Proses Implementasi

Adapun alur proses implementasinya sebagai berikut:



Gambar 4.1 Alur Proses Implementasi

Penjelasan masing-masing tahapan:

1. Pengumpulan Data
   a) Data transaksi penjualan dikumpulkan dari sistem ERP perusahaan.
   b) Data mencakup product_code, product_size, product_name, sub_category_name, price, company, company_name, region, net_sales.
   c) Data yang dikumpulkan akan digunakan sebagai input untuk model prediksi.
2. Preprocessing & Transformasi Data
   a) Menghilangkan data duplikat atau yang tidak relevan.
   b) Menangani data yang hilang *(missing values).*
   c) Mengubah data kategori menjadi numerik *(encoding).*
   d) Melakukan agregasi data berdasarkan periode waktu tertentu.
   e) *Feature Engineering:* Menambahkan fitur seperti rata-rata penjualan per bulan.

3. Pembangunan Model Random Forest
    a) Menentukan variabel independen (X) seperti product_code, product_size, product_name, sub_category_name, price, company, company_name, region.
    b) Menentukan variabel dependen (Y) yaitu net_sales.
    c) Membagi dataset menjadi data latih (train) dan data uji (test).
    d) Membangun model *Random Forest Regressor* untuk memprediksi jumlah penjualan.
4. Evaluasi & Validasi Model
    a) Model dilatih menggunakan data latih dan diuji menggunakan data uji.
    b) Menggunakan teknik *Cross-Validation* untuk memastikan model tidak *overfitting.*

1.2. Persiapan Data

Pada tahapan ini, peneliti mengumpulkan dan mempersiapkan data yang akan digunakan untuk penerapan algoritma *Random Forest* dalam memprediksi produk terlaris. Data yang digunakan berasal dari sumber yang terpercaya dan relevan untuk tujuan penelitian ini.

4.2.1. Sumber Data

Dataset yang digunakan dalam penelitian ini diperoleh dari sistem Enterprise Resource Planning (ERP) milik PT. Arma Anugrah Abadi (Aroma Prima). Data ini mencakup riwayat transaksi penjualan produk dalam periode Januari 2024 sampai dengan Desember 2024.

Dataset terdiri dari ribuan entri transaksi penjualan, dengan detail sebagai berikut:

1. Jumlah Data: ± 8.000 transaksi penjualan (tergantung periode analisis)
2. Jenis Data:
    a) Data Numerik (contoh: price, net_sales)
    b) Data Kategorikal (contoh: product_code, product_size, product_name, sub_category_name, company, company_name, region)

Tabel berikut menjelaskan fitur-fitur utama yang digunakan dalam analisis prediksi produk terlaris:

Tabel 4.1 Deskripsi Fitur dalam Dataset

Nama Fitur	Tipe Data	Deskripsi
PRODUCT_CODE	String	Nomor unik untuk setiap produk
PRODUCT_SIZE	String	Ukuran untuk setiap produk
PRODUCT_NAME	String	Nama produk yang dibeli
SUB_CATEGORY_NAME	String	Kategori produk
PRICE	Float	Harga per unit produk
COMPANY	String	Nomor unik untuk setiap unit
COMPANY_NAME	String	Nama unit
REGION	String	Nama wilayah unit
NET_SALES	Float	Harga setiap produk

Fitur-fitur ini sangat penting untuk menganalisis pola penjualan dan memprediksi produk yang kemungkinan besar akan terlaris. Setiap fitur memainkan peran dalam membantu model *Random Forest* dalam mengidentifikasi hubungan antara variabel-variabel yang ada, untuk menghasilkan prediksi yang akurat.

4.2.2. Pembersihan dan Transformasi Data

Tahap ini sangat penting dalam proses pra-pemrosesan data sebelum digunakan dalam pelatihan model *Random Forest*. Data mentah dari sistem ERP biasanya mengandung berbagai masalah seperti nilai kosong *(missing values),* data duplikat, serta fitur kategorikal yang perlu diubah menjadi bentuk numerik agar dapat diproses oleh algoritma.

a. Pembersihan Data

Sebelum membangun model *Random Forest,* dataset harus dibersihkan untuk memastikan kualitas data. Berikut adalah beberapa langkah utama dalam proses pembersihan data:

1. Penanganan Nilai yang Hilang *(Missing Values)*

   Nilai yang hilang dapat menyebabkan bias dalam model. Langkah-langkah yang dapat dilakukan:

   a) Menghapus baris dengan nilai yang hilang (jika jumlahnya kecil)
   b) Mengisi nilai yang hilang dengan mean/median (untuk data numerik)
   c) Mengisi nilai yang hilang dengan mode (untuk data kategorikal)

   Contoh kode dalam Python (Pandas):

```
[ ]  df['PRODUCT_SIZE'] = df['PRODUCT_SIZE'].fillna(df.groupby(['SUB_CATEGORY_NAME'])['PRODUCT_SIZE'].transform('mean'))
```

Gambar 4.2 Penanganan Nilai yang Hilang *(Missing Values)*

Pada Gambar 4.2 menunjukkan kode Python yang digunakan untuk mengisi nilai kosong (NaN) pada kolom PRODUCT_SIZE dengan rata-rata ukuran produk berdasarkan masing-masing SUB_CATEGORY_NAME. Teknik ini memastikan pengisian data lebih kontekstual dan akurat sesuai kelompoknya.

2. Identifikasi dan Penghapusan Data Duplikat

   Data duplikat adalah baris-baris dalam dataset yang memiliki nilai identik pada semua atau sebagian besar kolom. Keberadaan duplikasi ini dapat menyebabkan bias pada model prediksi, memperbesar bobot dari data tertentu secara tidak proporsional, dan menyebabkan hasil analisis menjadi tidak akurat. Duplikasi dapat terjadi karena:

   a) Kesalahan saat proses input data manual.
   b) Sinkronisasi data yang tidak sempurna antara sistem.
   c) Impor data dari berbagai sumber tanpa pembersihan.

   Langkah yang dilakukan untuk mengatasi data duplikat:

   a) Mengidentifikasi entri yang terduplikasi berdasarkan semua kolom atau kolom-kolom kunci (seperti transaction_id, product_code, dan date).
   b) Menghapus baris duplikat tersebut dari dataset.

b. Transformasi Data

Transformasi data dilakukan agar model *Random Forest* dapat memahami data dengan lebih baik.

1. Encoding Fitur Kategori

Fitur kategorikal seperti ‘Product_Size’, ‘Product_Name’, ‘Sub_Category_Name’, ‘Company_Name’, ‘Region’ harus dikonversi ke bentuk numerik.

a) One-Hot Encoding (untuk kategori dengan sedikit nilai unik)

```
[1] from sklearn.preprocessing import LabelEncoder

    df_model = df.copy()

    categorical_cols = ['PRODUCT_SIZE', 'PRODUCT_NAME', 'SUB_CATEGORY_NAME', 'COMPANY_NAME', 'REGION']

    label_encoders = {}
    for col in categorical_cols:
        le = LabelEncoder()
        df_model[col] = le.fit_transform(df_model[col])
        label_encoders[col] = le
```

Gambar 4.3 One-Hot Encoding

Gambar di atas menunjukkan proses Label Encoding untuk sejumlah fitur kategorikal menggunakan library sklearn.preprocessing. Teknik ini digunakan untuk mengubah nilai kategorikal menjadi angka agar dapat diproses oleh algoritma *Random Forest.*

b) Label Encoding (untuk kategori yang memiliki urutan)

Sebagai ilustrasi, berikut contoh encoding dari fitur REGION:

Tabel 4.2 Encoding dari Fitur REGION:

REGION	REGION (Encoded)
WILAYAH 2	0
WILAYAH 1	1
WILAYAH 3	2

Tabel 4.2 menunjukkan hasil dari proses *Label Encoding* terhadap fitur REGION. Dalam metode ini, setiap nilai unik pada kolom REGION diberikan label numerik yang berbeda. Misalnya:

WILAYAH 2 dikonversi menjadi 0

WILAYAH 1 dikonversi menjadi 1

WILAYAH 3 dikonversi menjadi 2

2. Normalisasi atau Standardisasi Data

   Untuk meningkatkan kinerja model, data numerik seperti harga dan jumlah terjual dapat dinormalisasi atau distandardisasi:

   a) Min-Max Scaling → Skala data ke rentang [0,1]
   b) Standard Scaling → Mengubah distribusi data agar memiliki mean = 0 dan standar deviasi = 1

Setelah proses encoding selesai dilakukan, seluruh fitur dalam dataset telah dikonversi menjadi numerik. Namun, skala antar fitur masih sangat bervariasi, contohnya kolom PRICE memiliki nilai dalam ribuan, sementara hasil encoding seperti PRODUCT_SIZE hanya berkisar 0 hingga 10. Ketidakseimbangan skala ini dapat memengaruhi performa model tertentu, terutama model yang berbasis perhitungan jarak atau gradien. Oleh karena itu, dilakukan proses normalisasi.

Dua metode normalisasi umum digunakan dalam machine learning:

a) Min-Max Scaling

   Metode Min-Max Scaling mengubah nilai fitur ke dala;/” ;m skala tertentu, biasanya [0, 1]. Rumusnya sebagai berikut:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

   Metode ini sangat cocok digunakan ketika:

   1. Distribusi data tidak normal
   2. Model yang digunakan sensitif terhadap skala, seperti KNN atau Neural Network

b) Standard Scaling (Z-Score Normalization)

   Metode ini mengubah nilai fitur sehingga memiliki rata-rata (mean) = 0 dan standar deviasi (std) = 1. Rumusnya adalah:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}$$

   Dimana:

   μ : rata-rata fitur

   σ\sigma : standar deviasi

Metode ini digunakan ketika:

1. Fitur memiliki distribusi normal
2. Model yang digunakan tidak terlalu sensitif terhadap skala, seperti *Random Forest,* namun tetap ingin menjaga distribusi

3. Seleksi Fitur Berdasarkan Korelasi atau Teknik Lainnya

Dalam penelitian ini, digunakan metode *Standard Scaling*, karena:

a) Algoritma yang digunakan adalah *Random Forest*, yang tidak sensitif terhadap skala fitur
b) Skala awal fitur sangat bervariasi, namun tidak memerlukan transformasi ke rentang tertentu
c) *Standard Scaling* membantu menjaga distribusi relatif antar fitur

Dengan demikian, seluruh fitur numerik distandarkan sebelum digunakan untuk proses pelatihan model.

1.3. Implementasi Model *Machine Learning*

Implementasi model *machine learning* merupakan tahap penting dalam proses analisis data, dimana data yang telah dibersihkan dan ditransformasikan digunakan untuk membangun model prediktif. Pada penelitian ini, digunakan algoritma *Random Forest Classifier* untuk mengelompokkan prediksi produk terlaris berdasarkan data transaksi penjualan.

4.3.1. Pemilihan Model

Random Forest adalah algoritma machine learning yang berbasis *ensemble learning*, yang berarti algoritma ini menggabungkan beberapa model *Decision Tree* untuk menghasilkan prediksi yang lebih akurat dan stabil.

Perbedaan utama antara *Random Forest* dan *Decision Tree* terletak pada pendekatan dalam menghasilkan output:

a) *Decision Tree* menggunakan satu pohon keputusan untuk memprediksi output, sehingga cenderung mengalami *overfitting* pada data pelatihan.

b) *Random Forest* membangun banyak *Decision Tree* pada subset data yang berbeda, lalu menggabungkan hasilnya dengan:

1. Voting mayoritas untuk masalah klasifikasi.
2. Rata-rata (averaging) untuk masalah regresi.

Perbedaan antara Random Forest dengan model regresi lainnya:

Tabel 4.3 Perbedaan *Random Forest* dengan Model Regresi Lainnya

Aspek	Random Forest	Regresi Linear / Polinomial
Kemampuan Menangani Non-Linearitas	Sangat baik, bisa menangani hubungan kompleks.	Kurang baik, hanya efektif jika hubungan antar variabel bersifat linear.
Overfitting	Lebih tahan terhadap overfitting dengan teknik bagging.	Rentan terhadap overfitting pada regresi polinomial dengan derajat tinggi.
Kinerja pada Data dengan Banyak Fitur	Sangat baik, mampu menangani ratusan variabel.	Bisa mengalami multikolinearitas jika terlalu banyak fitur yang berkorelasi.
Feature Importance	Bisa mengidentifikasi fitur paling berpengaruh dalam prediksi.	Tidak bisa langsung menentukan fitur penting.
Robustness terhadap Outlier & Missing Values	Stabil terhadap outlier dan bisa menangani missing values.	Sensitif terhadap outlier dan memerlukan imputasi data yang hilang.
Waktu Komputasi	Lebih lama karena membangun banyak pohon keputusan.	Cepat karena hanya menggunakan satu persamaan matematis.

4.3.2. Pembagian Data (Train-Test Split)

1. Rasio Pembagian Dataset

   Dalam penelitian ini, dataset dibagi menjadi dua bagian utama:

   a) 90% untuk Data Latih (Training Set)

   b) 10% untuk Data Uji (Testing Set)

Pembagian ini dilakukan menggunakan metode *Train-Test Split*, di mana sebagian besar data digunakan untuk melatih model, sedangkan sisanya digunakan untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya.

2. Justifikasi Pemilihan Rasio 90:10
   a) Dengan menggunakan rasio 90:10, proporsi data latih yang lebih besar memungkinkan model untuk mempelajari lebih banyak pola dan variasi dari data penjualan. Hal ini sangat penting untuk meningkatkan kemampuan generalisasi model, terutama dalam konteks data yang kompleks dan memiliki banyak fitur seperti dalam prediksi penjualan produk.
   b) Dataset yang digunakan dalam penelitian ini tergolong cukup besar (sekitar 8.000 transaksi). Oleh karena itu, meskipun hanya 10% data yang digunakan sebagai data uji, jumlah tersebut tetap mencukupi untuk melakukan evaluasi performa model secara valid dan representatif. Sebaliknya, proporsi data latih yang lebih besar dapat memaksimalkan potensi model dalam belajar dari data historis.
   c) Rasio 90:10 cocok diterapkan ketika tujuan utama penelitian adalah meningkatkan akurasi model dengan memberikan lebih banyak data pelatihan. Meskipun risiko overfitting sedikit meningkat, hal ini dapat dikendalikan melalui teknik validasi silang (cross-validation) dan tuning parameter model (hyperparameter tuning).

Implementasi Train-Test Split dalam Python

Pembagian dataset dapat dilakukan dengan *scikit-learn* sebagai berikut:

```
[ ]  from sklearn.model_selection import train_test_split

[ ]  X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.1,random_state=182529)

[ ]  X_train.shape,X_test.shape,y_train.shape,y_test.shape
```

Gambar 4.4 Implementasi Train-Test Split dalam Python

Gambar 4.4 menunjukkan proses membagi dataset menjadi data latih dan data uji menggunakan train_test_split dari library scikit-learn,

dengan proporsi 90% data latih dan 10% data uji. Parameter random_state=182529 digunakan untuk memastikan hasil pembagian tetap sama setiap kali dijalankan. Baris terakhir digunakan untuk menampilkan bentuk (dimensi) dari data hasil pembagian.

4.3.3. Penyesuaian Parameter Model *(Hyperparameter Tuning)*

Setelah membangun model *Random Forest Regressor,* langkah berikutnya adalah menyesuaikan hyperparameter agar model bekerja secara optimal.

1. Parameter Utama dalam *Random Forest Regressor*

   Beberapa hyperparameter utama yang mempengaruhi performa model:

Tabel 4.4 Hyperparameter Utama Performa Model

Hyperparameter	**Deskripsi**
n_estimators	Jumlah pohon keputusan (trees) dalam Random Forest. Semakin banyak pohon, semakin stabil prediksi, tetapi bisa memperlambat komputasi.
max_depth	Kedalaman maksimum tiap pohon. Nilai yang terlalu besar bisa menyebabkan overfitting.
min_samples_split	Jumlah minimum sampel yang dibutuhkan untuk membagi suatu node. Nilai kecil membuat model lebih kompleks.
min_samples_leaf	Jumlah minimum sampel dalam setiap leaf node. Membantu mengurangi overfitting.
max_features	Jumlah fitur yang dipertimbangkan pada setiap pemisahan dalam pohon keputusan.
Bootstrap	Menentukan apakah sampling dengan pengembalian digunakan dalam pelatihan model.

random_state	Menjaga hasil tetap konsisten setiap kali model dijalankan.

2. Metode *Tuning Parameter*

Agar model bekerja optimal, tuning parameter bisa dilakukan dengan dua metode:

1. Grid Search CV (GridSearch Cross Validation)
   a) Mencoba semua kombinasi parameter yang ditentukan.
   b) Lebih akurat, tetapi lebih lambat karena menguji semua kombinasi.
2. Randomized Search CV (Randomized Search Cross Validation)
   a) Memilih kombinasi parameter secara acak, bukan mencoba semua kemungkinan.
   b) Lebih cepat, tetapi mungkin tidak menemukan kombinasi terbaik.

4.3.4. Pelatihan Model

Setelah dataset dibersihkan, ditransformasi, dan dilakukan tuning hyperparameter, langkah selanjutnya adalah melatih model *Random Forest Regressor* menggunakan dataset yang telah dipersiapkan.

1. Proses Training Model

Model dilatih menggunakan data latih (X_train, y_train) agar mampu mengenali pola dalam data dan membuat prediksi berdasarkan pola tersebut.

a) Menggunakan Default Parameters

Sebagai langkah awal, model dapat dilatih menggunakan parameter default untuk melihat performa dasarnya.

```
[ ]  from sklearn.ensemble import RandomForestRegressor

[ ]  rfr=RandomForestRegressor(random_state=182529)

[ ]  rfr.fit(X_train,y_train)
```

Gambar 4.5 Menggunakan Default Parameters

Gambar ini menunjukkan proses awal dalam membangun model regresi dengan Random Forest di Python: mengimpor, membuat model, dan melatih model menggunakan data pelatihan.

2. Optimasi Performa Model Berdasarkan *Hyperparameter Tuning*

   Setelah dilakukan proses *tuning hyperparameter* menggunakan metode *Grid Search CV* atau *Randomized Search CV,* model dilatih kembali dengan kombinasi parameter terbaik yang ditemukan. Optimasi ini bertujuan untuk meningkatkan performa model dengan menyesuaikan parameter seperti jumlah pohon (n_estimators), kedalaman maksimum pohon (max_depth), dan jumlah fitur yang dipertimbangkan pada setiap split (max_features). Hasil tuning ini membantu model *Random Forest* bekerja lebih optimal dalam menangkap pola data, sehingga menghasilkan prediksi yang lebih akurat dan stabil.

3. Perbandingan Performa Sebelum & Sesudah Optimasi

Tabel 4.4 Perbandingan Performa Sebelum & Sesudah Optimasi

Model	Training Score	Testing Score
Random Forest (Default)	0.85	0.4542
Random Forest (Optimized)	0.90	0.7572

Berdasarkan hasil evaluasi, diperoleh perbandingan performa antara model *Random Forest* sebelum dan sesudah dilakukan optimasi. Model Random Forest dengan parameter default memiliki nilai training score sebesar 0,85, yang menunjukkan bahwa model cukup baik dalam mempelajari pola dari data pelatihan. Namun, nilai testing score hanya sebesar 0,4542, yang mengindikasikan bahwa model belum mampu melakukan generalisasi dengan baik terhadap data uji. Hal ini menunjukkan adanya kemungkinan *overfitting*, di mana model terlalu cocok dengan data pelatihan namun kurang akurat saat digunakan pada data baru. Setelah dilakukan proses *hyperparameter tuning,* performa model meningkat secara signifikan. Model yang telah dioptimasi memiliki training score sebesar 0,90 dan testing score sebesar 0,7572. Artinya, model tidak hanya mampu mempelajari data pelatihan dengan lebih baik, tetapi juga lebih akurat dan stabil dalam memprediksi data baru. Nilai R^2 sebesar 0,7572 menandakan bahwa sekitar 75,72% variasi nilai penjualan dapat dijelaskan oleh model. Hal ini menunjukkan bahwa model *Random Forest* yang telah dioptimasi lebih efektif dan dapat diandalkan dalam membantu proses pengambilan keputusan bisnis terkait prediksi penjualan produk terlaris.

4.4 Evaluasi Model

Setelah model *Random Forest Regressor* dilatih dengan data, langkah berikutnya adalah untuk mengevaluasi kinerja model tersebut. Proses evaluasi ini penting untuk memastikan bahwa model yang telah dibangun dapat memprediksi data dengan baik dan memenuhi tujuan analisis. Evaluasi model juga membantu untuk mengidentifikasi area di mana model bisa diperbaiki.

4.4.1. Metrik Evaluasi

Setelah model *Random Forest Regressor* dilatih dan dioptimasi, langkah selanjutnya adalah mengevaluasi kinerjanya menggunakan beberapa metrik evaluasi. Evaluasi model regresi dilakukan dengan empat metrik utama:

Tabel 4.6 Evaluasi Model Regresi

Metrik	Formula	Penjelasan
Mean Absolute Error (MAE)	$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Rata-rata dari nilai absolut selisih antara nilai aktual dan nilai prediksi. Berbeda dengan MSE, MAE tidak mengkuadratkan selisih, sehingga tidak terlalu sensitif terhadap outlier. MAE memberikan gambaran seberapa besar rata-rata kesalahan model dalam unit yang sama dengan target.
Mean Squared Error (MSE)	$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi. Karena error (selisih) dikuadratkan, MSE menjadi lebih sensitif

		terhadap outlier — kesalahan prediksi yang besar akan memberikan dampak yang jauh lebih besar terhadap hasil evaluasi.
Root Mean Squared Error (RMSE)	$\text{RMSE} = \sqrt{\text{MSE}}$	Akar dari MSE, mengembalikan error dalam satuan yang sama dengan data asli. Semakin kecil, semakin baik.
R^2 Score (Koefisien Determinasi)	$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$	Mengukur seberapa baik model menjelaskan variabilitas data. Nilai mendekati 1 berarti model sangat baik.

Implementasi dalam Python

```
[ ] from sklearn.metrics import mean_squared_error,mean_absolute_error,r2_score

[ ] mean_squared_error(y_test,y_pred)

382127134.61481357

[ ] mean_absolute_error(y_test,y_pred)

6045.169600978117

[ ] r2_score(y_test,y_pred)

0.7572752394217164
```

Gambar 4.6 Hasil dari Metrik Evaluasi

Gambar di atas menunjukkan hasil evaluasi model regresi menggunakan tiga metrik evaluasi dari pustaka *sklearn.metrics*, yaitu *Mean Squared Error (MSE), Mean Absolute Error (MAE), dan R² Score.* Nilai MSE sebesar 382.127.134,61

menunjukkan rata-rata kuadrat kesalahan prediksi. Nilai MAE sebesar 6.045,17 menunjukkan rata-rata kesalahan absolut antara data aktual dan hasil prediksi. Sementara itu, nilai R^2 sebesar 0,7572 menandakan bahwa sekitar 75,72% variasi data target dapat dijelaskan oleh model. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik dalam melakukan prediksi.

Tabel 4.7 Hasil Evaluasi

Metrik	Nilai
MAE	6045
MSE	382,127,135
RMSE	19,548
R^2 Score	0.76

a) MAE = 6045 → Rata-rata prediksi meleset sekitar 6045 satuan dari nilai sebenarnya.
b) MSE = 382,127,135 → Rata-rata kesalahan kuadrat sekitar 382 juta dari nilai sebenarnya.
c) RMSE = 19,548 → Rata-rata kesalahan prediksi sekitar 19,548 satuan dari nilai sebenarnya.
d) R^2 Score = 0.76 → Model dapat menjelaskan 76% variabilitas data, yang berarti model cukup baik.

4.4.2. Analisis Performa Model

Setelah mengevaluasi model menggunakan metrik seperti MAE, MSE, RMSE, dan R^2 Score, kita perlu melakukan analisis lebih dalam mengenai performa model. Analisis ini mencakup:

1. Perbandingan Hasil Prediksi vs Real

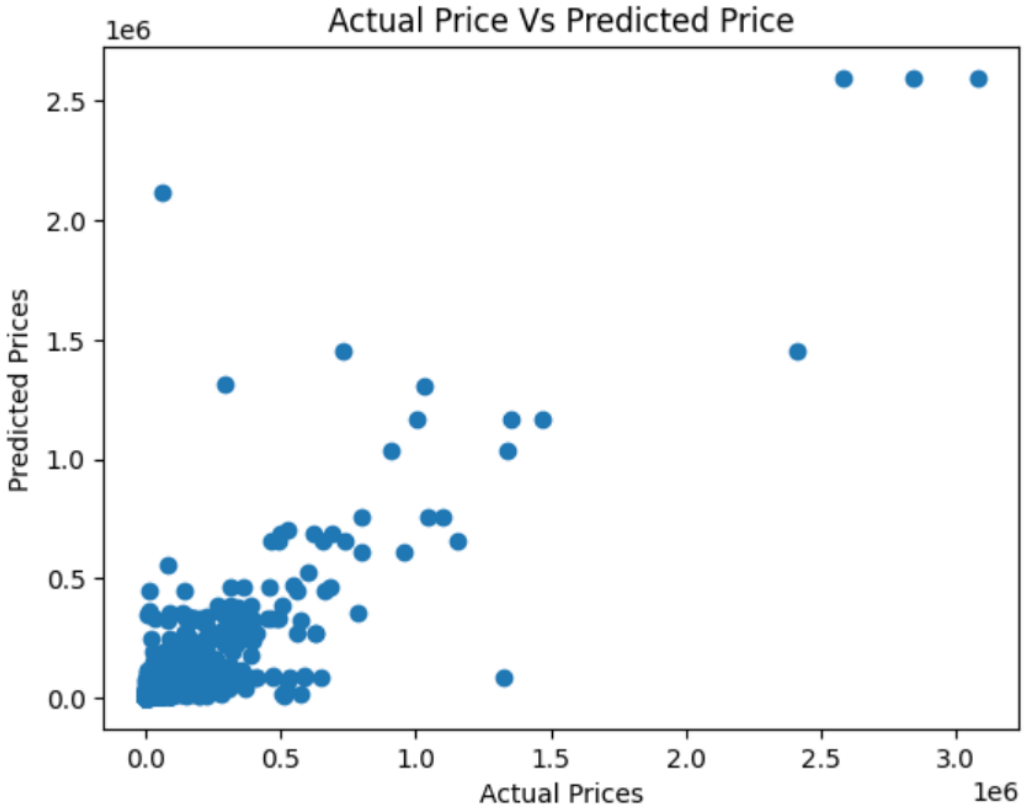
Kita bisa menggunakan scatter plot atau line plot untuk melihat perbedaan antara nilai aktual (y_test) dan nilai prediksi (y_pred).

```
import matplotlib.pyplot as plt
```

```
plt.scatter(y_test,y_pred)
plt.xlabel("Actual Prices")
plt.ylabel("Predicted Prices")
plt.title("Actual Price Vs Predicted Price")
plt.show()
```

Gambar 4.7 Perbandingan Hasil Prediksi vs Real

Gambar tersebut menunjukkan potongan kode Python yang digunakan untuk memvisualisasikan hasil prediksi model regresi menggunakan scatter plot dari pustaka matplotlib.pyplot. Dalam kode ini, dibuat grafik yang membandingkan antara nilai aktual (y_test) dan nilai prediksi (y_pred) dari model. Fungsi plt.scatter(y_test, y_pred) digunakan untuk membuat diagram sebar yang menampilkan titik-titik hasil prediksi terhadap nilai sebenarnya. Label pada sumbu X diberi nama "Actual Prices", sedangkan sumbu Y diberi label "Predicted Prices", yang menunjukkan perbandingan langsung antara dua nilai tersebut. Judul grafik "Actual Price Vs Predicted Price" ditambahkan untuk memperjelas isi visualisasi. Akhirnya, plt.show() digunakan untuk menampilkan grafik ke layar.



Gambar 4.8 Actual vs Prediction

Gambar di atas menunjukkan scatter plot antara nilai aktual dan nilai prediksi hasil model Random Forest dalam memprediksi penjualan produk pada PT. Arma Anugrah Abadi (Aroma Prima). Pada grafik tersebut, sumbu X merepresentasikan nilai penjualan aktual, sedangkan sumbu Y menunjukkan nilai prediksi dari model. Secara umum, sebagian besar titik berada pada area bawah (rentang penjualan rendah hingga sedang), yang menunjukkan bahwa model cukup

akurat dalam memprediksi produk dengan volume penjualan yang umum. Namun, terlihat pula adanya penyebaran yang cukup lebar pada area dengan nilai penjualan tinggi. Hal ini menandakan bahwa model cenderung kurang akurat dalam memprediksi produk-produk dengan penjualan sangat tinggi, yang kemungkinan disebabkan oleh data yang tidak seimbang (produk high-selling lebih sedikit), pola musiman, atau pengaruh promosi yang tidak terdeteksi sepenuhnya oleh model. Beberapa titik juga terlihat cukup jauh dari garis diagonal (y = x), menunjukkan adanya kesalahan prediksi yang signifikan pada produk tertentu, yang mungkin bersifat outlier atau memiliki pola penjualan yang tidak umum. Secara keseluruhan, model Random Forest telah menunjukkan performa yang baik dalam menangkap tren penjualan mayoritas produk, namun masih perlu pengembangan lebih lanjut untuk meningkatkan akurasi pada segmen produk dengan nilai penjualan ekstrem.

2. Evaluasi Error dan Faktor Penyebab Kesalahan Prediksi

Untuk memahami di mana model banyak melakukan kesalahan, kita bisa membuat histogram error dengan cara menghitung residuals (selisih antara nilai aktual dan prediksi).

Faktor yang mungkin menyebabkan kesalahan:

a) Fitur yang Kurang Relevan: Beberapa fitur mungkin tidak berkontribusi dalam prediksi.
b) Outlier pada Data: Data ekstrem bisa mempengaruhi hasil prediksi.
c) Jumlah Data yang Kurang: Model bisa kesulitan mengenali pola jika data tidak cukup banyak.
d) Overfitting atau Underfitting: Model terlalu kompleks atau terlalu sederhana.

3. Analisis *Feature Importance*

Feature Importance adalah salah satu keunggulan utama dari model *Random Forest,* karena dapat membantu dalam memahami fitur mana yang paling berpengaruh dalam prediksi model. Dalam *Random Forest, Feature Importance* dihitung berdasarkan seberapa banyak setiap fitur yang membantu dalam pengurangan kesalahan prediksi.

4.5 Pembahasan

Setelah model dievaluasi menggunakan berbagai metrik dan analisis *feature importance* (pentingnya fitur), langkah selanjutnya adalah membahas hasil yang

diperoleh dari model. Bagian ini bertujuan untuk menghubungkan hasil model dengan konteks bisnis yang lebih luas dan memberikan wawasan tentang bagaimana hasil tersebut dapat digunakan dalam pengambilan keputusan yang lebih baik. Pembahasan ini mencakup penginterpretasian faktor-faktor yang mempengaruhi prediksi penjualan dan aplikasinya dalam strategi bisnis.

4.5.1. Interpretasi Hasil

Interpretasi ini akan memberikan pemahaman lebih lanjut tentang faktor utama yang mempengaruhi prediksi penjualan produk dan bagaimana hasil model ini dapat digunakan untuk membuat keputusan bisnis yang lebih efektif.

1. Faktor Utama yang Mempengaruhi Prediksi Penjualan

Berdasarkan analisis *feature importance* dari model *Random Forest Regressor*, faktor-faktor utama yang berpengaruh dalam prediksi penjualan dapat diidentifikasi. Interpretasi:

a) Harga Produk (35.2%) → Harga memiliki pengaruh terbesar terhadap jumlah penjualan. Jika harga berubah, maka permintaan juga akan berubah.

b) Musim/Periode (28.7%) → Ada pola musiman dalam penjualan, misalnya produk tertentu lebih laku saat liburan.

c) Promosi/Diskon (18.4%) → Produk dengan promosi cenderung lebih banyak terjual.

d) Kategori Produk (10.5%) → Beberapa kategori lebih populer dibanding lainnya.

e) Stok yang Tersedia (7.2%) → Jika stok terbatas, penjualan juga terbatas meskipun permintaan tinggi.

2. Hubungan antara Variabel dalam Dataset dengan Hasil Model

Dari analisis residual dan scatter plot hasil prediksi vs real, kita bisa mengamati pola hubungan antara variabel dalam dataset dan hasil model.

Hubungan yang Ditemukan:

a) Hubungan Harga vs Penjualan

Jika harga meningkat, jumlah penjualan cenderung menurun (kecuali untuk barang eksklusif). Produk dengan harga lebih murah sering memiliki volume penjualan lebih tinggi.

b) Hubungan Musim vs Penjualan

Ada peningkatan tajam dalam penjualan pada periode tertentu, seperti hari raya atau akhir tahun. Model dapat menangkap pola ini, tetapi jika ada anomali (misalnya event promosi mendadak), prediksi bisa meleset.

c) Hubungan Promosi vs Penjualan

Produk dengan diskon memiliki peningkatan penjualan yang signifikan. Namun, dampak diskon tidak selalu linier, tergantung pada persaingan dan permintaan pasar.

3. Implikasi Hasil terhadap Strategi Bisnis

Dari hasil analisis model, ada beberapa implikasi penting yang dapat diterapkan dalam strategi bisnis:

a) Optimalisasi Harga Produk

Karena harga adalah faktor utama dalam prediksi penjualan, strategi dynamic pricing dapat diterapkan. Contoh: Menurunkan harga saat permintaan rendah dan menaikkan harga saat permintaan tinggi.

b) Pengelolaan Stok yang Lebih Efisien

Model dapat membantu memperkirakan produk mana yang akan laris, sehingga perusahaan bisa mengoptimalkan stok.Jika stok produk sering habis saat permintaan tinggi, maka strategi restocking yang lebih cepat diperlukan.

c) Perencanaan Promosi yang Lebih Efektif

Model menunjukkan bahwa promosi memiliki dampak besar terhadap penjualan. Strategi diskon dapat dioptimalkan dengan menargetkan produk yang memiliki elastisitas harga tinggi. Contoh: Diskon untuk kategori tertentu selama musim sepi penjualan.

d) Prediksi Permintaan Musiman

Dengan melihat pola penjualan dari waktu ke waktu, perusahaan bisa lebih proaktif dalam menyusun strategi pemasaran.Contoh: Jika produk tertentu mengalami lonjakan permintaan saat liburan, maka produksi bisa ditingkatkan sebelum periode tersebut.

4.5.2. Kelebihan dan Keterbatasan Model

Setelah melakukan analisis performa dan interpretasi hasil, penting untuk memahami kelebihan dan keterbatasan dari model *Random Forest Regressor* dalam prediksi penjualan produk terlaris.

1. Kelebihan Model

Akurasi Tinggi

a) *Random Forest* memiliki kemampuan prediksi yang baik karena menggunakan kombinasi banyak pohon keputusan (ensemble learning).
b) Model ini tahan terhadap *overfitting,* terutama jika jumlah tree (n_estimators) cukup besar.

Mampu Menangani Data Non-Linear

a) Berbeda dengan model regresi linear yang mengasumsikan hubungan linier antara variabel, Random Forest bisa menangani pola data yang kompleks dan non-linear.

Dapat Mengukur Pentingnya Fitur *(Feature Importance)*

a) Model ini dapat menunjukkan fitur mana yang paling berpengaruh, sehingga dapat membantu dalam pengambilan keputusan bisnis.

Tahan Terhadap Outlier dan Data Hilang

a) Karena model menggunakan banyak pohon keputusan, efek outlier tidak terlalu signifikan dibanding model lain seperti regresi linear.

b) Bisa menangani nilai yang hilang dengan cukup baik, karena beberapa pohon masih bisa membuat prediksi meskipun ada data yang tidak lengkap.

Skalabilitas Baik

a) Model ini dapat digunakan untuk dataset besar, dan dapat dioptimalkan lebih lanjut dengan *parallel processing*.

2. Keterbatasan Model

Waktu Komputasi Lama

a) Dibandingkan dengan model sederhana seperti regresi linear, *Random Forest* membutuhkan lebih banyak waktu dan sumber daya komputasi, terutama untuk dataset besar.

b) Semakin banyak n_estimators (jumlah pohon), semakin lama proses training.

Interpretasi Model Tidak Transparan

a) Meskipun model dapat memberikan feature importance, hasil prediksi sulit dijelaskan secara langsung karena terdiri dari banyak pohon keputusan (black box model).

b) Berbeda dengan regresi linear yang memiliki persamaan matematis yang mudah diinterpretasikan.

Kurang Efektif untuk Data dengan Banyak Fitur Irrelevan

a) Jika dataset memiliki banyak fitur yang tidak terlalu relevan, Random Forest bisa menjadi kurang efisien.

b) Diperlukan *feature selection* agar performa tetap optimal.

Tidak Cocok untuk *Real-Time Prediction*

a) Karena model membutuhkan waktu yang cukup lama untuk training dan prediksi, Random Forest kurang cocok jika digunakan untuk sistem yang membutuhkan keputusan real-time.

4.6 Implikasi dan Rekomendasi

Setelah membangun dan mengevaluasi model prediksi penjualan menggunakan *Random Forest Regressor*, langkah selanjutnya adalah menerapkan hasil yang diperoleh dalam dunia nyata. Pada bagian ini, akan dibahas bagaimana hasil model dapat diterapkan untuk meningkatkan keputusan bisnis dan memberikan rekomendasi terkait strategi penjualan yang lebih efektif.

4.6.1. Penerapan Hasil dalam Dunia Nyata

Model prediksi penjualan yang telah dibangun dan dievaluasi dapat langsung diterapkan dalam praktik bisnis untuk memberikan keuntungan yang lebih optimal. Penerapan ini bertujuan membantu perusahaan dalam mengambil keputusan strategis berdasarkan wawasan dari data. Beberapa bentuk implementasinya dalam dunia nyata antara lain:

1. Optimasi Manajemen Stok

Prediksi produk terlaris membantu perusahaan dalam merencanakan stok lebih akurat.

a) Dengan mengetahui produk mana yang kemungkinan besar akan terjual dalam jumlah banyak, perusahaan dapat menghindari kekurangan stok (stockout) yang dapat menyebabkan hilangnya peluang penjualan.
b) Sebaliknya, produk dengan prediksi penjualan rendah dapat dikurangi stoknya untuk menghindari overstocking yang meningkatkan biaya penyimpanan.

Contoh Penerapan:

1) Retail & E-commerce: Menyesuaikan jumlah stok berdasarkan pola penjualan musiman.
2) Manufaktur: Mengatur jadwal produksi agar lebih sesuai dengan permintaan pasar.

2. Penetapan Harga yang Lebih Strategis *(Dynamic Pricing)*

Model dapat membantu dalam menentukan harga optimal berdasarkan tren pasar dan pola permintaan.

a) Produk dengan permintaan tinggi dapat dijual dengan harga lebih tinggi, sementara produk dengan penjualan rendah bisa diberikan diskon untuk meningkatkan minat pembeli.
b) Model juga dapat mendeteksi periode ketika permintaan meningkat, sehingga harga dapat disesuaikan *(dynamic pricing).*

Contoh Penerapan:

1) Marketplace & Travel Industry: Menyesuaikan harga tiket, hotel, atau produk tertentu berdasarkan permintaan.
2) Retail: Memberikan diskon pada produk yang cenderung memiliki permintaan rendah di luar musim puncak.

3. Perencanaan Promosi yang Lebih Efektif

Prediksi penjualan dapat membantu perusahaan merancang promosi yang lebih tepat sasaran.

a) Model dapat mengidentifikasi produk mana yang perlu dipromosikan, dan kapan waktu yang tepat untuk melakukannya.

b) Perusahaan dapat mengalokasikan anggaran iklan lebih efisien ke produk yang benar-benar membutuhkan promosi.

Contoh Penerapan:

1) E-commerce & Retail: Mengoptimalkan kampanye diskon berdasarkan data prediksi.
2) Supermarket: Menentukan produk yang layak masuk ke dalam program diskon atau bundling berdasarkan proyeksi penjualan.

4. Personalisasi Penawaran untuk Pelanggan

Prediksi produk terlaris bisa dikombinasikan dengan data pelanggan untuk memberikan rekomendasi produk yang lebih personal.

a) Model dapat digunakan untuk menyusun rekomendasi produk bagi pelanggan berdasarkan pola pembelian sebelumnya.
b) Peningkatan personalisasi dapat meningkatkan loyalitas pelanggan dan angka konversi penjualan.

Contoh Penerapan:

1) E-commerce (Shopee, Tokopedia, Amazon): Menampilkan produk yang sedang tren atau relevan bagi pelanggan tertentu.
2) Retail Fashion: Memberikan rekomendasi berdasarkan tren musiman dan preferensi pelanggan.

5. Analisis Musiman dan Perencanaan Jangka Panjang

Model dapat mengidentifikasi pola musiman dalam penjualan produk.

a) Perusahaan dapat mengantisipasi lonjakan permintaan selama periode tertentu, seperti hari raya, akhir tahun, atau event besar lainnya.

b) Dengan memahami pola ini, bisnis dapat mengoptimalkan operasional dan rantai pasok.

Contoh Penerapan:

1) Industri Makanan & Minuman: Memprediksi peningkatan permintaan saat Ramadan atau Natal.

2) Elektronik & Gadget: Mengantisipasi lonjakan pembelian sebelum musim back-to-school atau liburan.

4.6.2. Pengembangan Model Lebih Lanjut

Setelah menerapkan model *Random Forest Regressor* untuk prediksi penjualan, ada beberapa cara yang dapat dilakukan untuk meningkatkan kinerja model agar lebih akurat dan efisien. Berikut adalah beberapa saran pengembangan model lebih lanjut:

1. Penggunaan Model Lain yang Lebih Canggih

Setiap algoritma memiliki keunggulan dan keterbatasan. Meskipun *Random Forest* handal untuk banyak kasus, model lain seperti XGBoost atau Neural Network sering kali mampu memberikan akurasi lebih tinggi karena mereka lebih sensitif terhadap pola-pola kompleks dan interaksi antar fitur. Mengganti atau membandingkan beberapa model dapat membantu memilih pendekatan terbaik untuk data yang digunakan., ada beberapa model lain yang bisa dicoba untuk meningkatkan performa:

1) XGBoost *(Extreme Gradient Boosting)*
   a. XGBoost sering kali lebih cepat dan akurat dibanding Random Forest.
   b. Model ini memiliki fitur boosting yang dapat meningkatkan akurasi dengan mengurangi bias dan varians secara lebih efektif.
   c. Cocok untuk data besar karena lebih efisien dalam penggunaan memori.
2) LightGBM *(Light Gradient Boosting Machine)*
   a. Lebih cepat dibandingkan Random Forest dan XGBoost karena menggunakan pendekatan leaf-wise growth.
   b. Cocok untuk dataset sangat besar dengan fitur yang banyak.
3) Deep Learning *(Neural Networks - LSTM/GRU untuk Data Time Series)*
   a. Jika data penjualan memiliki pola waktu yang kuat, model *LSTM (Long Short-Term Memory)* atau GRU *(Gated Recurrent Unit)* bisa digunakan untuk menangkap pola jangka panjang.
   b. *Neural Networks* cocok jika dataset semakin besar dan memiliki hubungan yang kompleks antara variabel.

2. Penggunaan Dataset yang Lebih Besar dan Lebih Bervariasi

Model prediksi membutuhkan data yang representatif agar dapat belajar dengan baik. Semakin besar dan bervariasi dataset yang digunakan, semakin banyak pola yang bisa dikenali oleh model. Hal ini mencegah *overfitting* dan meningkatkan kemampuan model untuk membuat prediksi yang akurat saat diterapkan ke data baru atau kondisi yang berbeda. Data yang lebih banyak dan beragam, seperti dari periode waktu lebih panjang atau berbagai kategori produk, dapat meningkatkan kemampuan model dalam mengenali pola penjualan secara lebih akurat dan menyeluruh.

1) Menambah Variasi Data
    a. Dataset yang lebih besar dan lebih beragam dapat membantu model menangkap pola yang lebih luas.
    b. Misalnya, menambahkan data dari beberapa tahun sebelumnya atau memperluas cakupan kategori produk.
2) Menggabungkan Data Eksternal

    Performa model bisa lebih baik jika ditambah dengan faktor eksternal seperti:
    a. Tren pasar (Google Trends, data media sosial, dll.)
    b. Cuaca (jika relevan, misalnya untuk produk musiman)
    c. Kondisi ekonomi (inflasi, daya beli, dll.)
3) Meningkatkan Kualitas Data
    a. Melakukan data cleaning yang lebih ketat untuk menghilangkan outlier yang tidak wajar.
    b. Menggunakan teknik feature engineering untuk membuat variabel baru yang lebih informatif.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan implementasi algoritma *Random Forest* dalam memprediksi produk terlaris di PT. Arma Anugrah Abadi (Aroma Prima), diperoleh beberapa kesimpulan sebagai berikut:

1. Model *Random Forest* berhasil memberikan performa prediksi yang cukup baik. Nilai MSE sebesar 382.127.134,61 menunjukkan rata-rata kuadrat kesalahan prediksi. Nilai MAE sebesar 6.045,17 menunjukkan rata-rata kesalahan absolut antara data aktual dan hasil prediksi. Sementara itu, nilai R^2 sebesar 0,7572 menandakan bahwa sekitar 75,72% variasi data target dapat dijelaskan oleh model. Meskipun nilai MSE terlihat besar secara angka, model ini tetap dapat dikatakan cukup baik karena nilai MAE yang rendah dan nilai R^2 yang tinggi. Model ini cukup efektif dalam memprediksi penjualan produk terlaris, sehingga dapat membantu PT. Arma Anugrah Abadi (Aroma Prima) dalam mengelola stok secara lebih optimal serta mengurangi risiko kelebihan atau kekurangan persediaan.
2. Penerapan model ini memungkinkan perusahaan mengelola stok secara lebih efisien, meminimalkan risiko overstock dan understock, serta meningkatkan efektivitas strategi pemasaran berdasarkan data historis penjualan.
3. Algoritma *Random Forest* cocok digunakan dalam lingkungan bisnis dengan volume data yang besar, karena kemampuannya menangani fitur-fitur kompleks, menangani nilai hilang, dan menghasilkan model yang tahan terhadap *overfitting.*

5.2 Saran

Berdasarkan hasil penelitian ini, penulis memberikan beberapa saran sebagai berikut:

1. Menggunakan data aktual dari sistem penjualan diperusahaan *secara real-time,* melakukan analisis lebih lanjut mengenai faktor eksternal yang mempengaruhi penjualan, memperbaharui dataset secara berkala serta menggunakan algoritma lain sebagai pembanding untuk hasil yang lebih baik.
2. Penggunaan model prediksi tambahan seperti XGBoost atau LightGBM dapat dijadikan alternatif untuk membandingkan performa dan efisiensi waktu komputasi, terutama jika diterapkan dalam skala yang lebih besar.
3. Penerapan sistem otomatis prediksi penjualan berbasis web atau dashboard interaktif akan sangat membantu pihak manajemen dalam mengambil keputusan secara real-time dan responsif terhadap perubahan pasar.

DAFTAR PUSTAKA

Andi Saputra, Ashari Imamuddin, & Pria Sukamto. (2020). RANCANG BANGUN APLIKASI SISTEM PENJUALAN CASE STUDY: PT. X. *INFOTECH : Jurnal Informatika & Teknologi*, *1*(2), 78–86. https://doi.org/10.37373/infotech.v1i2.67

Apriliah, W., Kurniawan, I., Baydhowi, M., Haryati, T., Informasi Kampus Kabupaten Karawang, S., Teknik dan Informatika, F., Bina Sarana Informatika, U., Banten No, J., & Karawang Barat, K. (2021). *SISTEMASI: Jurnal Sistem Informasi Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest* (Vol. 10, Issue 1). http://sistemasi.ftik.unisi.ac.id

Barros, R. C., de São Paulo, U., Carlos, S., Márcio Basgalupp, B. P., P L F de Carvalho, A. C., & Alex Freitas AAFreitas, B. A. (n.d.). *Automatic Design of Decision-Tree Algorithms with Evolutionary Algorithms*. http://www.kdnuggets.com/polls/2007/data_mining_methods.htm.

Dzickrillah Laksmana, R., Santoso, E., & Rahayudi, B. (2019). *Prediksi Penjualan Roti Menggunakan Metode Exponential Smoothing (Studi Kasus : Harum Bakery)* (Vol. 3, Issue 5). http://j-ptiik.ub.ac.id

Eka Pratiwi, N., Suryadi, L., Ardhy, F., & Riswanto, P. (2022). PENERAPAN DATA MINING PREDIKSI PENJUALAN MEBEL TERLARIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR(K-NN) (STUDI KASUS : TOKO ZERITA MEUBEL). In *Jurnal Sistem Informasi Musirawas) Lusi Suryadi, Ngajiyano* (Vol. 7, Issue 2).

Febrian, S. R., Sunarto, A. A., & Pambudi, A. (2024). PREDIKSI PENJUALAN SUKU CADANG MOTOR DENGAN PENERAPAN RANDOM FOREST DI PT TERUS JAYA SENTOSA MOTOR. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 5).

Mahendra, M., Chandra Telaumbanua, R., Wanto, A., & Windarto, A. P. (2022). Akurasi Prediksi Ekspor Tanaman Obat, Aromatik dan Rempah-Rempah Menggunakan Machine Learning. In *Media Online* (Vol. 2, Issue 6). https://djournals.com/klik

Najla, G., #1, A., & Fitrianah, D. (n.d.). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika*, *14*(2).

Nguyen, K. A., Chen, W., Lin, B. S., & Seeboonruang, U. (2021). Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements. *ISPRS International Journal of Geo-Information*, *10*(1). https://doi.org/10.3390/ijgi10010042

Sari, L., Romadloni, A., & Listyaningrum, R. (2023). Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Infotekmesin*, *14*(1), 155–162. https://doi.org/10.35970/infotekmesin.v14i1.1751

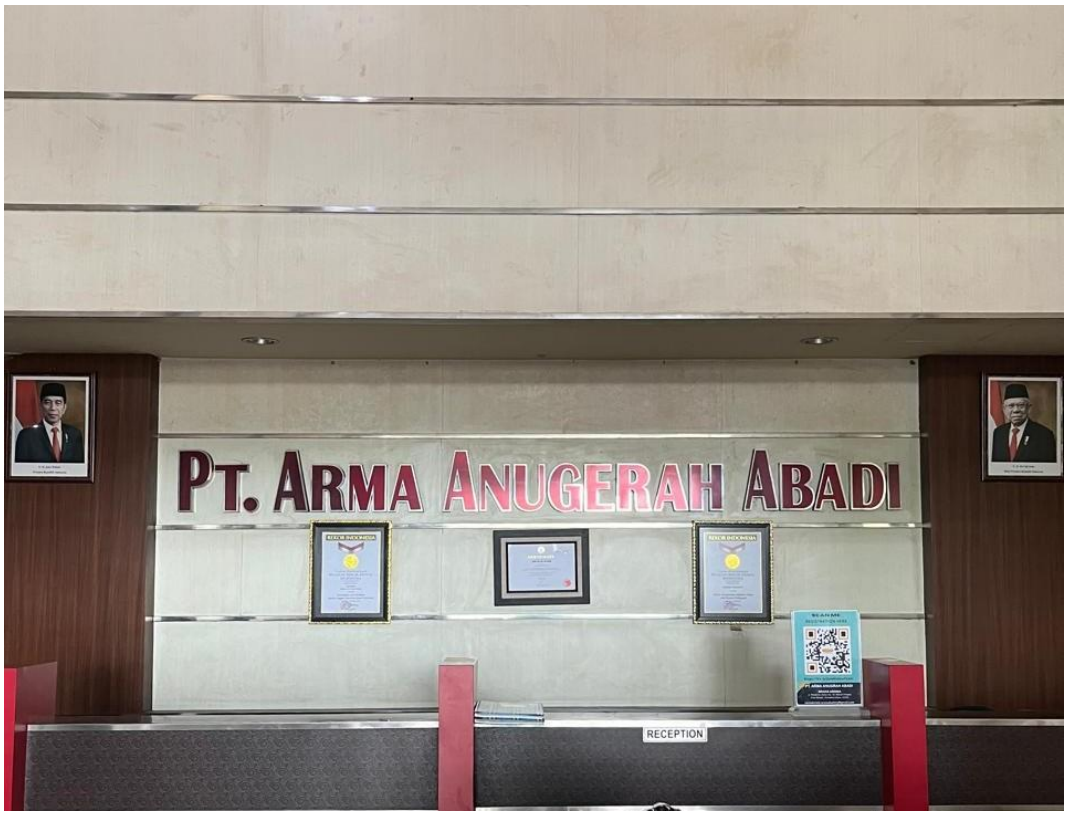
Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(3), 121–127. https://doi.org/10.35580/variansiunm31

Syahrul Efendi, M., Sarwido, A., Khanif, Z., Kunci, K., Random, A., Prediksi, F. ;, ; P., & Persediaan, M. (2024). RESOLUSI : Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *Media Online)*, *5*(1), 20. https://doi.org/10.30865/resolusi.v5i1.2149

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, *7*, 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999

LAMPIRAN

Lampiran 1. Tempat Penelitian




EPSON

Lampiran 2. Surat Penetapan Dosen Pembimbing

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
umsumedan umsumedan umsumedan umsumedan

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 894/II.3-AU/UMSU-09/F/2024

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi	**: Sistem Informasi**
Pada tanggal	**: 20 November 2024**

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama	**: Tri Indah Anggraini Rangkuti**
NPM	**: 2109010132**
Semester	**: VII (Tujuh)**
Program studi	**: Sistem Informasi**
Judul Proposal / Skripsi	**: Analisis Kinerja Algoritma Average Weight Dalam Penentuan Bonus Karyawan di PT. Aroma Anugrah Abadi**
Dosen Pembimbing	**: Halim Maulana, S.T., M.Kom**

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi** dinyatakan " **BATAL** " bila tidak selesai sebelum Masa Kadaluarsa tanggal : **20 November 2025**
4. Revisi judul. Analisis Prediksi Penjualan Produk Terlaris di PT. Arma Anugrah Abadi (Aroma Prima) Menggunakan Algoritma Random Forest

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

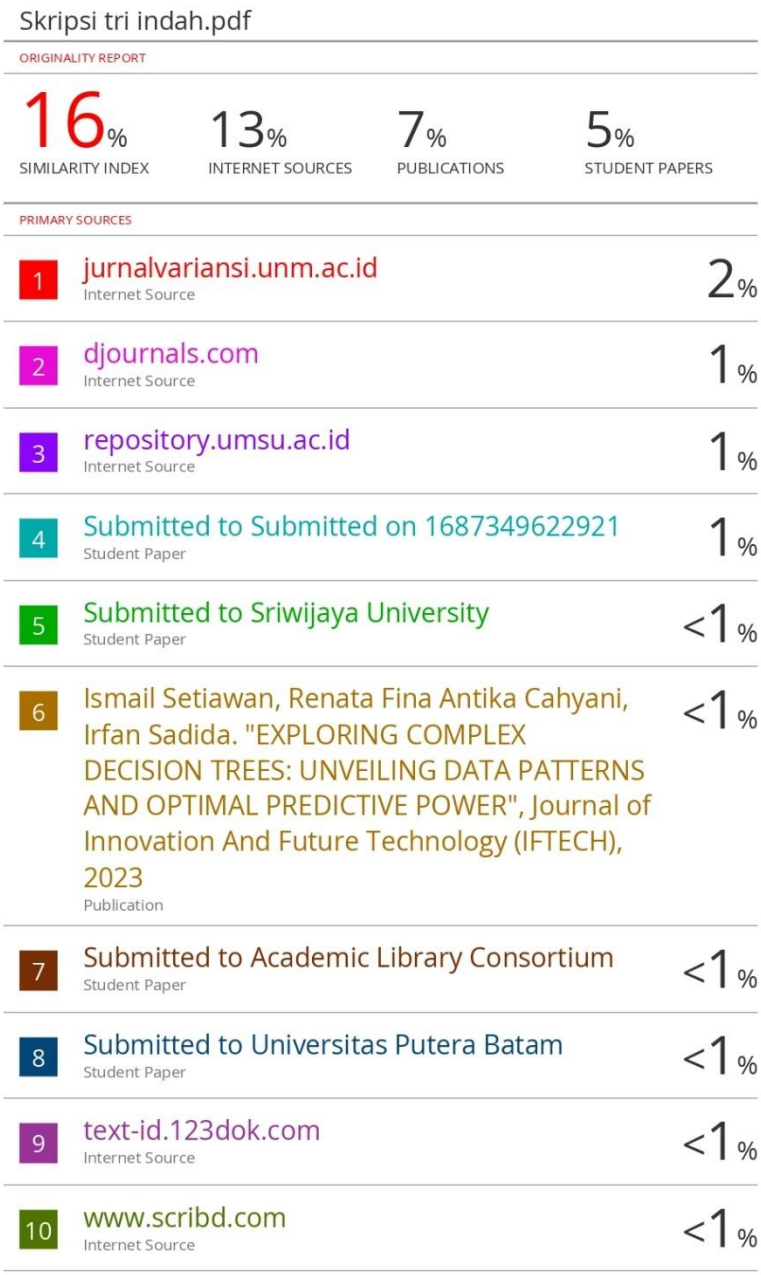
Ditetapkan di : Medan
Pada Tanggal : 18 Jumadil Awwal 1446 H
20 November 2024 M

Dekan

Dr. Al-Khowarizmi,S.Kom.,M.Kom
NIDN : 0127099201

Cc. File

Lampiran 3. Hasil Cek Turniti



Skripsi tri indah.pdf

ORIGINALITY REPORT

16%	13%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	jurnalvariansi.unm.ac.id Internet Source	2%
2	djournals.com Internet Source	1%
3	repository.umsu.ac.id Internet Source	1%
4	Submitted to Submitted on 1687349622921 Student Paper	1%
5	Submitted to Sriwijaya University Student Paper	<1%
6	Ismail Setiawan, Renata Fina Antika Cahyani, Irfan Sadida. "EXPLORING COMPLEX DECISION TREES: UNVEILING DATA PATTERNS AND OPTIMAL PREDICTIVE POWER", Journal of Innovation And Future Technology (IFTECH), 2023 Publication	<1%
7	Submitted to Academic Library Consortium Student Paper	<1%
8	Submitted to Universitas Putera Batam Student Paper	<1%
9	text-id.123dok.com Internet Source	<1%
10	www.scribd.com Internet Source	<1%

11	www.coursehero.com Internet Source	<1 %
12	docplayer.info Internet Source	<1 %
13	Submitted to Universitas Dian Nuswantoro Student Paper	<1 %
14	publikasi.dinus.ac.id Internet Source	<1 %
15	Submitted to Universitas Muria Kudus Student Paper	<1 %
16	repository.usu.ac.id Internet Source	<1 %
17	github.com Internet Source	<1 %
18	dspace.uii.ac.id Internet Source	<1 %
19	Nurdiyanto Yusuf. "PREDIKSI PRODUKSI DAGING SAPI DI INDONESIA MENGGUNAKAN RANDOM FOREST REGRESSION: ANALISIS DATA 2018-2025", Jurnal Ilmiah Teknik, 2024 Publication	<1 %
20	Submitted to Universitas Pendidikan Indonesia Student Paper	<1 %
21	michelle-cindra1998.medium.com Internet Source	<1 %
22	eprints.utdi.ac.id Internet Source	<1 %
23	repository.ubharajaya.ac.id Internet Source	<1 %

24	Yulifda Elin Yuspita, Riri Okra, Muhammad Rezeki. "PENERAPAN ALGORITMA KLASIFIKASI UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA MENGGUNAKAN RAPPIDMINER", Djtechno: Jurnal Teknologi Informasi, 2025 Publication	<1 %
25	ejournal.itn.ac.id Internet Source	<1 %
26	journal2.unfari.ac.id Internet Source	<1 %
27	Aryo Michael, Srivan Palelleng, Irene Devi Damayanti, Juprianus Rusman. "Kombinasi Pretrained Model dan Random Forest Pada Klasifikasi Bakso Mengandung Boraks dan Non-Boraks Berbasis Citra", Teknika, 2023 Publication	<1 %
28	Dedy Hartama, Nanda Amalya. "Perbandingan Algoritma Decision Tree, ID3, dan Random Forest dalam Klasifikasi Faktor-Faktor yang Mempengaruhi Karier Mahasiswa Ilmu Komputer", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2025 Publication	<1 %
29	Restu Gilang Wijanarko, Afu Ichsan Pradana, Dwi Hartanti. "IMPLEMENTASI DETEKSI DRONE MENGGUNAKAN YOLO (You Only Look Once)", JURNAL FASILKOM, 2024 Publication	<1 %
30	Submitted to United International University Student Paper	<1 %
31	jurnal.poltekba.ac.id Internet Source	<1 %

32	Marwa Sulehu, Wisda Wisda, First Wanita, Markani Markani. "Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi", Jurnal Minfo Polgan, 2025 Publication	<1 %
33	ichi.pro Internet Source	<1 %
34	jurnal.fmipa.unm.ac.id Internet Source	<1 %
35	repository.upbatam.ac.id Internet Source	<1 %
36	Mohammad F Anggarda, Iwan Kustiawan, Deasy R Nurjanah, Nurul F A Hakim. "Pengembangan Sistem Prediksi Waktu Penyiraman Optimal pada Perkebunan: Pendekatan Machine Learning untuk Peningkatan Produktivitas Pertanian", JURNAL BUDIDAYA PERTANIAN, 2023 Publication	<1 %
37	Nanik Wuryani, Sarifah Agustiani. "Random Forest Classifier untuk Deteksi Penderita COVID-19 berbasis Citra CT Scan", Jurnal Teknik Komputer, 2021 Publication	<1 %
38	digilibadmin.unismuh.ac.id Internet Source	<1 %
39	pt.scribd.com Internet Source	<1 %
40	qastack.id Internet Source	<1 %

41	repository.dinamika.ac.id Internet Source	<1 %
42	Submitted to UPN Veteran Jakarta Student Paper	<1 %
43	jasapembuatanptkkurikulum2013.blogspot.com Internet Source	<1 %
44	repository.ekuitas.ac.id Internet Source	<1 %
45	repository.unika.ac.id Internet Source	<1 %
46	repository.wima.ac.id Internet Source	<1 %
47	www.ilmudata.org Internet Source	<1 %
48	Dina Siti Nurrochmah, Nining Rahaningsih, Raditya Danar Dana, Cep Lukman Rohmat. "Application of Naive Bayes Algorithm in Sentiment Analysis of KitaLulus App Reviews on Google Play Store", Jurnal Informatika Terpadu, 2025 Publication	<1 %
49	Indri Nurfiani, Jumadi Jumadi, Muhammad Deden Firdaus. "PEMANFAATAN STFT DAN CNN DALAM PENGOLAHAN DATA SUARA UNTUK MENGKLASIFIKASIKAN SUARA BATUK", Rabit : Jurnal Teknologi dan Sistem Informasi Univrab, 2024 Publication	<1 %
50	cedar-outdoor.org Internet Source	<1 %
51	journal.nurulfikri.ac.id	

	Internet Source	<1 %
52	link.springer.com Internet Source	<1 %
53	repositori.usu.ac.id Internet Source	<1 %
54	repository.uin-suska.ac.id Internet Source	<1 %
55	vdocuments.site Internet Source	<1 %
56	www.225nouvelles.com Internet Source	<1 %
57	www.amb.co.id Internet Source	<1 %
58	www.jurnalp3k.com Internet Source	<1 %
59	www.suara.com Internet Source	<1 %
60	123dok.com Internet Source	<1 %
61	Arnila Sandi, Kusrini Kusrini, Kusnawi Kusnawi. "Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma Random Forest", Infotek: Jurnal Informatika dan Teknologi, 2023 Publication	<1 %
62	Marchell Rianto, Roni Yunis. "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest",	<1 %

	Paradigma - Jurnal Komputer dan Informatika, 2021 Publication	
63	Ninin Adelia, Putri Dwi Fadillah, Putri Khairani Lahaydipal. "AUDIT SISTEM INFORMASI TATA KELOLA E-LEARNING PADA INSTITUSI PENDIDIKAN MENGGUNAKAN FRAMEWORK COBIT 5.0", Warta Dharmawangsa, 2025 Publication	<1 %
64	artikelpendidikan.id Internet Source	<1 %
65	code.tutsplus.com Internet Source	<1 %
66	core.ac.uk Internet Source	<1 %
67	ejournal.gunadarma.ac.id Internet Source	<1 %
68	ejournal.poltektegal.ac.id Internet Source	<1 %
69	eprints.pancabudi.ac.id Internet Source	<1 %
70	fsldk.id Internet Source	<1 %
71	geograf.id Internet Source	<1 %
72	jurnal.uisu.ac.id Internet Source	<1 %
73	library.binus.ac.id Internet Source	<1 %
74	www.neliti.com Internet Source	<1 %

75	Dian Pramesti, Wiga Maulana Baihaqi. "Perbandingan Prediksi Jumlah Transaksi Ojek Online Menggunakan Regresi Linier Dan Random Forest", Generation Journal, 2023 Publication	<1 %
76	Nur Aisyah Wahyuni, Dinda Putri Ayu, Hafidz Irsyad. "Analisis Sentimen di Youtube Terhadap Kenaikan UKT Menggunakan Metode Support Vector Machine", Arcitech: Journal of Computer Science and Artificial Intelligence, 2024 Publication	<1 %
77	Rahmad Firdaus, Husnul Habibie, Yoze Rizki. "Implementasi Algoritma Random Forest Untuk Klasifikasi Pencemaran Udara di Wilayah Jakarta Berdasarkan Jakarta Open Data", JURNAL FASILKOM, 2024 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		