EXPLORE
Jurnal Sistem Informasi dan Telematika
(Telekomunikasi, Multimedia dan Informatika)

Implementasi Algoritma Random Forest untuk Klasifikasi Keberhasilan Penjualan Produk Kopi pada UMKM di Lampung

Muhammad Reza Romahdoni[1], Rahma Diani Sormin[2] , Levina Rachmawati Putri[3], Rudi Kurniawan[4], Sevi Andriasari[5]

[1,5] Jurusan Rekayasa Perangkat Lunak, Institut Teknologi Dan Bisnis Diniyyah Lampung, Pesawaran, Indonesia
[2,3] Jurusan Kewirausahaan, Institut Teknologi Dan Bisnis Diniyyah Lampung, Pesawaran, Indonesia
[4] Jurusan Manajemen Ritel, Institut Teknologi Dan Bisnis Diniyyah Lampung, Pesawaran, Indonesia
m.rezarhomadoni@gmail.com, rahmadianisormin@gmail.com, levinarp25@gmail.com, rudi.kurniawan2506@gmail.com, andriasari.sevii@gmail.com

Abstract- Lampung Province is recognized as one of the largest coffee production centers in Indonesia. Small and Medium Enterprises (SMEs) serve as the primary drivers of coffee downstreaming in the region, with a significant growth of local coffee shops and processed coffee producers. However, in their daily operations, they often still rely on intuition or traditional methods to predict market demand and the success of their product sales. This research classifies the sales success of coffee products in Lampung SMEs using the Random Forest algorithm. The research stages include problem identification, data collection, data pre-processing, Random Forest implementation, and model evaluation. This systematic method is expected to produce an accurate classification model. Based on the research conducted, it can be concluded that the implementation of the Random Forest algorithm provides an effective and accurate solution for classifying the sales success of coffee products in Lampung SMEs. With an accuracy rate of 94% and an Area Under the Curve (AUC) value of 0.984, the developed model demonstrates excellent generalization capability in recognizing complex transaction patterns. This model is expected to be developed into a web or mobile-based application that is more user-friendly, allowing the benefits of data science to be accessed and applied directly by SMEs in their daily operational activities without requiring deep technical expertise.

Keywords: Machine Learning, Random Forest, Classification, Sales, MSMEs

Abstrak- Provinsi Lampung dikenal sebagai salah satu sentra produksi kopi terbesar di Indonesia. Sektor Usaha Mikro, Kecil, dan Menengah menjadi penggerak utama hilirisasi kopi di daerah ini, dengan banyaknya kedai kopi dan produsen kopi olahan lokal yang tumbuh signifikan. namun dalam operasional kesehariannya, mereka sering kali masih mengandalkan intuisi atau metode tradisional dalam memprediksi permintaan pasar dan keberhasilan penjualan produk merekaPenelitian ini mengklasifikasikan keberhasilan penjualan produk kopi UMKM Lampung menggunakan algoritma Random Forest. Tahapan penelitian mencakup identifikasi masalah, pengumpulan data, pra-pemrosesan data, implementasi Random Forest, serta evaluasi model. Metode sistematis ini diharapkan menghasilkan model klasifikasi akuratBerdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma Random Forest berhasil memberikan solusi yang efektif dan akurat dalam mengklasifikasikan keberhasilan penjualan produk kopi pada UMKM di Lampung. Dengan tingkat akurasi mencapai 94% dan nilai Area Under the Curve (AUC) sebesar 0,984, model yang dibangun terbukti memiliki kemampuan generalisasi yang sangat baik dalam mengenali pola transaksi yang kompleks.Model ini diharapkan dapat dikembangkan menjadi sebuah sistem aplikasi berbasis web atau mobile yang lebih ramah pengguna (user-friendly), sehingga manfaat data science ini dapat diakses dan diaplikasikan langsung oleh pelaku UMKM dalam kegiatan operasional sehari-hari tanpa memerlukan keahlian teknis yang mendalam.

Kata Kunci: Machine Learning, Random Forest, Klasifikasi, Penjualan, UMKM

1. Pendahuluan

Di era digital yang berkembang pesat, pemanfaatan *data science* dan *machine learning* telah menjadi tulang punggung bagi efisiensi operasional di berbagai sektor industri global [1]. Kemampuan untuk mengolah data transaksi yang masif menjadi informasi prediktif memungkinkan perusahaan untuk mengambil keputusan berbasis data yang lebih akurat, responsif, dan terukur [2]. Pendekatan analitik ini tidak lagi terbatas pada perusahaan berskala besar, tetapi mulai merambah ke sektor bisnis yang lebih kecil guna menghadapi ketidakpastian pasar yang semakin dinamis [3].

Provinsi Lampung dikenal sebagai salah satu sentra produksi kopi terbesar di Indonesia, yang didominasi oleh komoditas kopi Robusta [4]. Sektor Usaha Mikro, Kecil, dan Menengah (UMKM) menjadi penggerak utama hilirisasi kopi di daerah ini, dengan banyaknya kedai kopi dan produsen kopi olahan lokal yang tumbuh signifikan [5]. UMKM kopi di Lampung memegang peranan krusial dalam ekonomi daerah, namun dalam operasional kesehariannya, mereka sering kali masih mengandalkan intuisi atau metode tradisional dalam memprediksi permintaan pasar dan keberhasilan penjualan produk mereka.

Ketergantungan pada intuisi tersebut kerap menimbulkan berbagai permasalahan, seperti ketidakakuratan dalam estimasi stok dan ketidakefektifan strategi promosi. Banyak pelaku UMKM kopi di Lampung kesulitan mengidentifikasi variabel-variabel kunci yang mempengaruhi klasifikasi keberhasilan penjualan, seperti jenis kopi, *grade* kopi, metode proses, harga per kilogram, *channel* penjualan, status sertifikasi, hingga tingkat kepuasan pelanggan (*rating*). Kompleksitas interaksi antara faktor-faktor tersebutdi mana status keberhasilan penjualan ditentukan oleh kombinasi variabel yang dinamis sering kali tidak tertangkap oleh metode tradisional, sehingga potensi keuntungan yang seharusnya bisa dioptimalkan menjadi hilang atau tidak maksimal.

Beberapa penelitian terdahulu yang membahas topik serupa diantaranya di lakukan oleh [6]. Hasil penelitian menunjukkan bahwa metode Naive Bayes berhasil mencapai akurasi sebesar 75%, serta dapat mengidentifikasi pola penjualan musiman yang sangat penting untuk perencanaan strategi pemasaran dan pengadaan stok. Penelitian lain di lakukan oleh [6]. Data dikumpulkan dari transaksi penjualan dan digunakan untuk mengklasifikasikan produk berdasarkan algoritma C4.5, yang diimplementasikan melalui alat *RapidMiner*. Hasil penelitian menunjukkan bahwa algoritma C4.5 efektif dalam menentukan keakuratan klasifikasi produk terlaris dan tidak laris, sehingga membantu meminimalkan masalah overstock dan kekurangan stok di gudang, yang dapat meningkatkan efisiensi operasional dan daya saing bisnis. Selanjutnya penelitian di lakukan oleh [7]. Mendefinisikan algoritma klasifikasi penambangan data yang akurat untuk memprediksi keberhasilan telemarketing berdasarkan eksperimen 2010. dalam pemasaran produk jasa perbankan, hasil proses evaluasi algoritma ini ditentukan dengan *cross-validation*, *Confusion Matrix, ROC curve dan T-test*. Algoritma regresi logistik lebih akurat dengan akurasi 92,32 % dan nilai *AUC* 0,962, sehingga algoritma yang digunakan termasuk dalam kelompok klasifikasi yang baik. Kemudian penelitian di lakukan oleh [8]. Hasil pengujian menunjukkan akurasi sebesar 76%, dengan *precision* 0,79, *recall* 0,98, dan *F1-score* 0,87 pada kategori laris. Temuan ini membuktikan bahwa Naive Bayes mampu memberikan hasil prediksi yang cukup andal untuk kategori mayoritas, namun kinerjanya menurun pada kategori minoritas akibat ketidakseimbangan distribusi data. Penelitian ini menyimpulkan bahwa algoritma Naive Bayes dapat digunakan sebagai alat bantu pengambilan keputusan dalam manajemen stok dan strategi penjualan UMKM, serta merekomendasikan penerapan teknik penyeimbangan data atau eksplorasi algoritma lain pada penelitian berikutnya untuk meningkatkan performa di semua kategori.

Sebagian besar studi literatur terdahulu mengenai optimasi penjualan UMKM di Indonesia masih bertumpu pada analisis deskriptif konvensional atau penggunaan algoritma klasifikasi tunggal yang sederhana seperti Naïve Bayes dan C4.5, yang kerap rentan terhadap *overfitting* serta kurang optimal dalam menangani kompleksitas interaksi fitur non-linier pada data transaksi produk agrikultur lokal. Di sisi lain, para pelaku UMKM kopi di Lampung secara faktual masih mengandalkan intuisi subjektif dalam manajemen stok dan strategi pemasaran.

Kebaruan dari penelitian ini terletak pada pembangunan model prediktif terintegrasi berbasis *ensemble learning* menggunakan algoritma *Random Forest* yang dikhususkan untuk komoditas hilirisasi kopi lokal unggulan di Provinsi Lampung. Penelitian ini tidak hanya menyajikan model klasifikasi dengan tingkat akurasi tinggi, tetapi juga memetakan kontribusi fitur spesifik (*feature importance*) serta merumuskan implikasi manajerial berbasis data (*data-driven decision making*) yang dapat diadopsi secara langsung oleh ekosistem UMKM daerah.

Mengingat peranan UMKM sebagai pilar ekonomi Lampung, urgensi untuk melakukan transformasi digital dalam manajemen penjualan menjadi sangat mendesak [9]. Tanpa adanya sistem prediksi yang andal, pelaku UMKM akan terus terjebak dalam siklus "coba-coba" (*trial and error*) yang berisiko tinggi secara finansial [10]. Oleh karena itu, diperlukan sebuah pendekatan sistematis yang mampu memetakan pola transaksi kompleks menjadi wawasan bisnis yang mudah dipahami, sehingga keberlangsungan usaha dapat lebih terjamin di tengah persaingan pasar yang semakin ketat

Sebagai solusi alternatif, penerapan algoritma *machine learning* menjadi langkah yang sangat relevan. Di antara berbagai algoritma yang tersedia, *Random Forest* menawarkan keunggulan dalam hal akurasi, kemampuannya menangani data kategorikal yang kompleks, serta ketahanannya terhadap *overfitting* [11].

EXPLORE: ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: http://dx.doi.org/10.36448/jsit.v17i1.5000

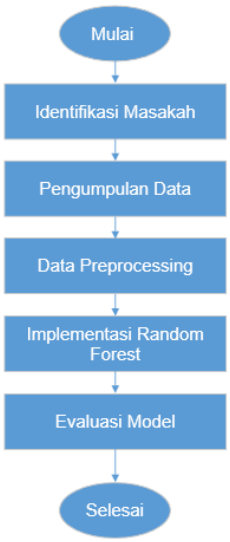
Dengan memanfaatkan algoritma ini, hubungan antara variabel input seperti karakteristik produk dan strategi pemasaran dengan status keberhasilan penjualan sebagai variabel target dapat diklasifikasikan dengan presisi. Hal ini memungkinkan pelaku UMKM untuk mengetahui lebih awal probabilitas kesuksesan sebuah produk atau strategi tertentu sebelum diterapkan secara nyata.

Pemilihan algoritma *Random Forest* untuk mengklasifikasikan keberhasilan penjualan produk kopi didasarkan pada beberapa landasan teoretis yang kuat terkait karakteristik data transaksional UMKM. Pertama, data penjualan produk agrikultur sering kali melibatkan interaksi antar variabel yang kompleks dan bersifat non-linier (seperti kombinasi antara varietas kopi, metode pengolahan pascapanen, harga jual, dan tingkat kepuasan konsumen). Secara teoretis, *Random Forest* yang berbasis *ensemble learning* dari sekumpulan pohon keputusan (*decision trees*) mampu mempartisi ruang fitur secara rekursif ke dalam wilayah-wilayah keputusan non-linier yang sangat fleksibel. Kedua, pohon keputusan tunggal sangat rentan terhadap *overfitting*, terutama pada ukuran dataset UMKM yang terbatas. Berdasarkan teori *bootstrap aggregating (bagging)* dan metode subruang acak (*random subspace method*) yang dipelopori oleh Leo Breiman, *Random Forest* mengatasi hal ini dengan melatih setiap pohon secara independen pada sampel data acak serta memilih subset fitur secara acak pada setiap percabangan simpul (*node split*). Hal ini menghasilkan pohon-pohon yang saling terdekorasi (*decorrelated*), sehingga mampu mereduksi varians secara signifikan tanpa meningkatkan bias serta memberikan kemampuan generalisasi yang sangat baik terhadap data transaksi baru. Ketiga, struktur hierarkis *Random Forest* sangat andal dalam memproses data tipe campuran (*mixed data types*), baik variabel numerik (seperti harga dan *rating*) maupun kategorikal (seperti jenis dan grade kopi) yang menjadi karakteristik utama data operasional UMKM kopi.

Berdasarkan paparan tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma *Random Forest* guna mengklasifikasikan keberhasilan penjualan produk kopi pada UMKM di Lampung berdasarkan variabel-variabel tersebut. Melalui penelitian ini, diharapkan dapat dihasilkan sebuah model klasifikasi yang mampu memberikan *insight* strategis bagi para pelaku usaha kopi, sehingga mereka dapat mengoptimalkan efektivitas penjualan, meningkatkan efisiensi operasional, dan pada akhirnya memperkuat posisi daya saing UMKM kopi lokal di pasar regional maupun nasional.

2. Metodologi

Tahapan penelitian yang digunakan untuk mengklasifikasikan keberhasilan penjualan produk kopi pada UMKM di Lampung menggunakan algoritma Random Forest dapat di lihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

A. Identifikasi

Identifikasi masalah dalam penelitian ini bahwa praktik operasional UMKM kopi di Lampung saat ini masih sangat bergantung pada intuisi dan metode tradisional, sehingga pelaku usaha belum mampu memanfaatkan data historis sebagai basis pengambilan keputusan yang objektif. Terdapat kompleksitas yang belum terpetakan dalam memahami faktor-faktor penentu keberhasilan penjualan, di mana pelaku usaha kesulitan mengidentifikasi kontribusi variabel-variabel kunci seperti jenis kopi, *grade* kopi, metode proses, harga per kilogram, *channel* penjualan, status sertifikasi, hingga *rating* kepuasan pelanggan terhadap status keberhasilan produk di pasar. Keterbatasan dalam menganalisis interaksi antar variabel tersebut mengakibatkan strategi pemasaran dan manajemen stok menjadi tidak tepat sasaran dan cenderung bersifat *trial and error*, yang secara langsung berdampak pada ketidakefektifan operasional dan penurunan potensi keuntungan. Selain itu, belum tersedianya model klasifikasi yang sistematis untuk memproses data transaksi tersebut menyebabkan pelaku UMKM sulit untuk memprediksi probabilitas keberhasilan penjualan secara akurat. Kondisi ini menempatkan UMKM kopi lokal dalam posisi yang rentan terhadap dinamika pasar, sehingga diperlukan sebuah model berbasis *machine learning* untuk memberikan *insight* strategis guna meningkatkan efisiensi operasional dan memperkuat daya saing produk di pasar yang semakin kompetitif..

B. Pengumpulan Data

Penelitian ini menggunakan dataset transaksi penjualan produk kopi pada UMKM di Provinsi Lampung dengan total 100 instansi (data transaksi) dan 8 variabel/fitur tanpa missing values. Rincian variabel fitur dalam dataset tersebut meliputi: jenis kopi, grade kopi, metode proses, harga per kg, channel penjualan, status sertifikasi, rating kepuasan dan status keberhasilan sebagai variable target Data dikumpulkan melalui tiga metode

utama untuk memastikan validitas dan akurasi model klasifikasi yang akan dibangun, yaitu:

1) Wawancara

Teknik wawancara dilakukan secara terstruktur kepada para pemilik atau pengelola UMKM kopi di Lampung yang menjadi subjek penelitian. Tujuan utama dari wawancara ini adalah untuk memahami proses bisnis yang berjalan, menggali informasi mengenai variabel-variabel yang memengaruhi penjualan (jenis kopi, grade kopi, metode proses, harga per kg, channel penjualan, status sertifikasi, rating kepuasan), serta mengonfirmasi parameter yang dianggap sebagai "keberhasilan" bagi pelaku usaha.

2) Observasi

Observasi dilakukan dengan cara mengamati secara langsung alur operasional dan perilaku penjualan produk kopi di lokasi UMKM. Proses ini melibatkan pemantauan terhadap bagaimana transaksi terjadi, bagaimana produk dipasarkan melalui berbagai *channel*, serta bagaimana tanggapan pelanggan yang direpresentasikan melalui *rating* kepuasan. Observasi langsung ini sangat krusial untuk memvalidasi data historis yang diperoleh, memastikan kesesuaian antara catatan administratif dengan realitas di lapangan.

3) Studi Literatur

Studi literatur dilakukan dengan menelusuri berbagai referensi ilmiah, seperti buku, jurnal penelitian terdahulu, artikel teknis, dan dokumen terkait *machine learning* serta industri kopi. Fokus utama dari studi literatur ini adalah untuk memperdalam pemahaman mengenai algoritma *Random Forest*, metrik evaluasi model, dan standar industri dalam klasifikasi data penjualan.

C. Data Preprocessing

Tahapan *data preprocessing* dalam penelitian ini merupakan proses krusial untuk memastikan bahwa data transaksi yang terkumpul memiliki kualitas yang bersih, konsisten, dan representatif sebelum diolah lebih lanjut oleh algoritma *Random Forest* [12][13]. Proses ini diawali dengan pembersihan data (*data cleaning*), di mana setiap *missing values* ditangani melalui metode imputasi yang sesuai seperti penggunaan nilai modus untuk variabel kategorikal atau median untuk variabel numerik serta pembersihan *outlier* untuk memastikan tidak ada data anomali yang mendistorsi pola model. Berdasarkan karakteristik dataset transaksi UMKM kopi yang berjumlah 100 instansi dengan 8 variabel fitur tanpa missing values, tidak diterapkan metode statistik parametrik spesifik (seperti *Z-score* atau *Interquartile Range* / *IQR*) untuk deteksi *outlier*. Pembersihan anomali dilakukan secara prosedural pada tahap pembersihan awal data transaksi untuk memastikan integritas data sebelum diproses ke dalam model *Random Forest* Berdasarkan pemeriksaan awal terhadap keseluruhan dataset transaksi penjualan produk kopi, tidak terdapat nilai yang hilang (*zero missing values*) yang diidentifikasi sebelum tahap preprocessing, di mana seluruh 100 instansi data terisi secara lengkap pada 8 atribut variabel yang digunakan. Setelah data dinyatakan bersih, dilakukan tahap transformasi data melalui proses *encoding*, yaitu mengubah variabel kategorikal seperti jenis kopi, metode proses, dan *channel* penjualan ke dalam format numerik agar dapat diproses oleh sistem. Meskipun *Random Forest* cukup tangguh, proses *scaling* juga diterapkan pada variabel numerik seperti harga per kilogram untuk mengoptimalkan efisiensi komputasi model. Berkaitan dengan penerapan *feature scaling* pada variabel numerik seperti harga per kilogram, meskipun secara teoretis algoritma Random Forest berbasis pohon keputusan bersifat *scale-invariant* (tidak memerlukan normalisasi karena proses percabangan didasarkan pada ambang batas peringkat data), langkah ini tetap diselaraskan di dalam kerangka kerja *pipeline* perangkat lunak *Orange Data Mining* untuk menjaga konsistensi pemrosesan data, tanpa mengubah distribusi asli dari fitur numerik yang digunakan. Tahap selanjutnya adalah seleksi fitur (*feature selection*), di mana variabel yang paling relevan dengan status keberhasilan diidentifikasi guna meningkatkan akurasi dan efektivitas prediksi. Sebagai langkah akhir dari *preprocessing*, dilakukan pembagian dataset (*data splitting*) dengan rasio yang proporsional, seperti 80% untuk data pelatihan (*training set*) dan 20% untuk data pengujian (*testing set*). Pembagian ini dilakukan secara objektif untuk mencegah terjadinya *overfitting*, sehingga model yang dihasilkan benar-benar mampu melakukan generalisasi dan memberikan prediksi yang akurat terhadap data transaksi baru di masa depan.

Tabel 1. Deskripsi Variabel Prediktor dan Tipe Data

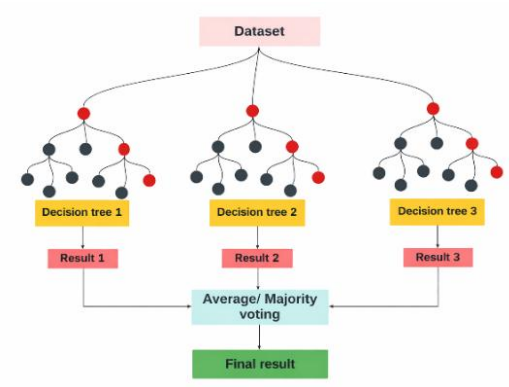
Nama Variabel (*Feature*)	Deskripsi Variabel	Tipe Data (*Data Type*)	Peran/ Model
jenis_kopi	Varietas kopi lokal (Arabika, *House Blend*, Robusta)	Kategorikal	Variabel (Fitur)
grade_kopi	Tingkat mutu kualitas biji kopi	Kategorikal	Variabel (Fitur)
metode_proses	Metode pengolahan pascapanen (*Full Wash*, *Honey*, *Natural*)	Kategorikal	Variabel (Fitur)
harga_per_kg	Harga jual produk kopi per kilogram	Numerik	Variabel (Fitur)
channel_penjualan	Saluran distribusi pemasaran (*Online*, *Offline*, Keduanya)	Kategorikal	Variabel (Fitur)

EXPLORE: ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: http://dx.doi.org/10.36448/jsit.v17i1.5000

status_sertifikasi	Status kepemilikan sertifikasi (Bersertifikat, Belum Bersertifikat)	Kategorikal	Variabel (Fitur)
rating_kepuasan	Tingkat kepuasan konsumen berdasarkan ulasan	Numerik	Variabel (Fitur)
status_keberhasilan	Status pencapaian penjualan produk (Berhasil, Gagal)	Kategorikal	Variabel Target (*Target*)

D. Implementasi Random Forest

Algoritma *Random Forest* merupakan salah satu metode *ensemble learning* berbasis pohon keputusan (*decision tree*) yang bekerja dengan cara membangun banyak pohon secara independen selama proses pelatihan dan menggabungkan hasil prediksinya melalui mekanisme *majority voting* untuk klasifikasi [14]:



Gambar 2. *Random Forest*

Rumus algoritma *Random Forest* dapat dilihat pada persamaan 1 [15].

$$\hat{f}(x) = \frac{1}{M}\sum_{j=1}^{M} h_j(x) \quad (1)$$

Keterangan:
F(x) : output dari *Random Forest*
J : jumlah pohon dalam ensemble
h_j (x) : output dari pohon ke - (j).

1) Jumlah Pohon (*Number of Tree*)

Parameter ini menentukan total pohon keputusan yang akan dibentuk di dalam *forest*. Semakin banyak jumlah pohon yang digunakan, model cenderung menjadi lebih stabil dan akurat karena keputusan akhir diambil berdasarkan agregasi suara dari berbagai sudut pandang sampel data.

2) Jumlah Atribut Pada Setiap Pemisah (*Number of attributes considered at each split)*

Parameter ini mengatur batasan jumlah fitur acak yang dievaluasi pada setiap simpul (*node*) ketika pohon melakukan percabangan. Dengan memilih subset fitur secara acak pada setiap pemisahan, model dapat memastikan adanya keragaman (*decorrelation*) antar pohon keputusan, sehingga mencegah dominasi oleh satu fitur tertentu saja.

3) Kontrol Pertumbuhan (*Growth Control*)

Pengaturan ini mencakup pembatasan kedalaman maksimum individu pohon (*limit depth of individual trees*) serta penentuan jumlah sampel minimum pada suatu subset sebelum percabangan dihentikan (*do not split subsets smaller than*). Pengaturan kontrol pertumbuhan ini sangat krusial untuk menjaga agar model tidak terlalu kompleks yang dapat memicu *overfitting* terhadap data latih, sehingga model memiliki kemampuan generalisasi (*generalization capability*) yang sangat baik saat diuji menggunakan data baru.

Implementasi algoritma *Random Forest* dalam penelitian ini dilakukan melalui serangkaian tahapan sistematis berbasis *visual programming* menggunakan *software Orange Data Mining* untuk membangun model klasifikasi yang handal [15]. Pemilihan perangkat lunak *Orange Data Mining* sebagai platform implementasi model klasifikasi didasarkan pada beberapa keunggulan strategis dibandingkan platform *machine learning* berbasis teks atau pemrograman murni. Pertama, Orange menyediakan antarmuka visual berbasis komponen (*visual programming / widget-based workflow*) yang mempermudah visualisasi alur data secara transparan dari tahap *preprocessing*, pemodelan, hingga evaluasi. Kedua, platform ini menawarkan fleksibilitas tinggi dalam pengujian interaktif melalui widget terintegrasi seperti *Test and Score, Confusion Matrix*, dan *Rank* tanpa memerlukan penulisan kode sintaks yang rumit, sehingga mempercepat proses validasi eksperimen. Ketiga, Orange sangat efisien untuk pengembangan model prototipe dan analisis data eksploratif pada dataset skala kecil hingga menengah seperti data operasional UMKM, dengan tetap menyediakan pustaka inti berbasis Python di latar belakang yang menjamin keandalan komputasi matematis yang setara dengan kerangka kerja pemrograman tradisional. Proses dimulai dengan memuat *dataset* melalui *widget* File, yang kemudian dilanjutkan dengan tahap preprocessing melalui *widget* seperti *Select Columns* untuk menetapkan variabel input dan variabel target. Setelah data siap, dilakukan proses pelatihan (*training*) dengan membagi *dataset* menggunakan *widget* Data Sampler (rasio 80:20), lalu menghubungkannya ke *widget Random Forest* untuk membangun sekumpulan *decision tree* yang dilatih secara independen. Di dalam *widget* ini, dilakukan optimasi model dengan menyesuaikan parameter seperti *n_estimators* dan *max_depth* secara interaktif. Selanjutnya, model diuji menggunakan *widget* Test and Score untuk menghasilkan metrik evaluasi seperti *Accuracy, Precision, Recall,* dan *F1-Score*, yang kemudian divalidasi melalui *Confusion Matrix* guna melihat

EXPLORE: ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: http://dx.doi.org/10.36448/jsit.v17i1.5000

distribusi prediksi yang akurat [16]. Sebagai bagian penting dari analisis strategis, *widget* Rank digunakan untuk mengidentifikasi *feature importance*, sehingga dapat diketahui variabel mana yang paling berpengaruh terhadap keberhasilan penjualan.

Tabel 2. Feature Importance

Nama Fitur (Variabel)	Skor Kontribusi (Importance)	Presentase
Rating Kepuasan	0.42	42 %
Harga per Kg	0.28	28 %
Channel Penjualan	0.15	15 %
Metode Proses	0.08	8 %
Jenis Kopi	0.04	4 %
Grade Kopi	0.02	2 %
Status Sertifikasi	0.01	1 %

E. Evaluasi Model

Tahapan evaluasi dimulai dengan pengujian model menggunakan data uji (*testing set*) yang telah dipisahkan sebelumnya guna memastikan kemampuan generalisasi model terhadap data baru. Dalam *widget Test and Score*, dihasilkan metrik evaluasi utama seperti *Accuracy* untuk melihat tingkat keberhasilan klasifikasi secara keseluruhan, serta *Precision* dan *Recall* untuk mengukur kualitas prediksi pada setiap kelas (berhasil atau tidak). Selain itu, metrik *F1-Score* digunakan sebagai penyeimbang antara *precision* dan *recall* untuk memberikan gambaran yang lebih komprehensif, terutama jika terdapat ketidakseimbangan data. Selanjutnya, *Confusion Matrix* digunakan untuk menganalisis distribusi kesalahan prediksi, di mana model akan memperlihatkan jumlah *True Positive* (prediksi berhasil yang tepat), *True Negative* (prediksi tidak berhasil yang tepat), serta *False Positive* dan *False Negative* sebagai indikator adanya kesalahan klasifikasi. Dengan melakukan evaluasi mendalam melalui tahapan ini, peneliti dapat memastikan apakah model telah mencapai standar akurasi yang diharapkan atau memerlukan penyesuaian parameter lebih lanjut guna meningkatkan performa prediksi bagi operasional UMKM kopi di Lampung

1) Akurasi (*Accuracy*)

Akurasi mengindikasikan seberapa efektif model dalam mengelompokkan data dengan benar secara keseluruhan [17]. Nilai ini diperoleh dengan menghitung rasio antara jumlah prediksi yang tepat terhadap total data yang digunakan. Persentase akurasi Persentase akurasi dalam penelitian ini dihitung menggunakan rumus berikut. dalam penelitian ini dihitung menggunakan rumus berikut. menghitung rasio antara jumlah prediksi yang tepat terhadap total data yang digunakan. Persentase akurasi dalam penelitian ini dihitung menggunakan rumus berikut.

$$Accurary = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

2) Presisi (*Precision*)

Presisi (*Positive Predictive Value*) digunakan untuk menilai kinerja sistem dengan menghitung data yang diklasifikasikan dengan benar dan data yang salah diklasifikasikan. Data yang terklasifikasikan dengan benar menjadi acuan untuk memperoleh nilai presisi, dengan membaginya dengan hasil prediksi

False Positive, yaitu data yang terprediksi tidak tepat. Persentase presisi dalam penelitian ini dihitung menggunakan rumus berikut.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3) *Recall*

Recall digunakan untuk mengukur kemampuan sistem dalam mengidentifikasi kembali seluruh instansi data yang benar-benar bernilai positif (seperti transaksi yang berstatus "Berhasil") dari keseluruhan total data positif aktual yang ada di dalam dataset. Nilai *recall* diperoleh dengan cara membandingkan jumlah prediksi yang benar-benar positif (*True Positive*) terhadap total keseluruhan data aktual yang seharusnya positif, yang merupakan penjumlahan dari *True Positive* (TP) dan *False Negative* (FN)..

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

4) *F1 Score*

F1-Score adalah nilai yang membandingkan nilai rata-rata antara presisi dan *Recall* [13]. Rumus yang digunakan untuk menghitung nilai F1-Score adalah sebagai berikut

$$\text{F1-Score} = 2\ x \frac{Recall\ x\ Precision}{Recall+Precision} \quad (5)$$

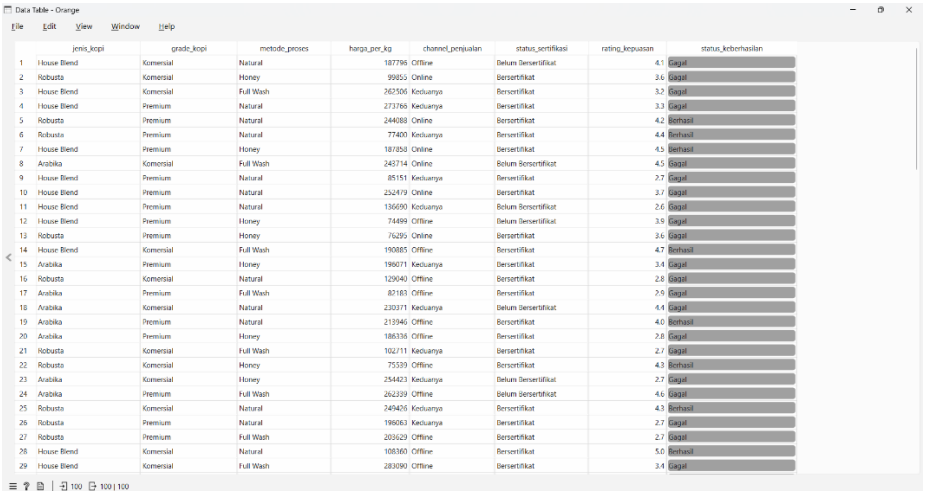
3. Hasil dan Pembahasan

A. Dataset

Penelitian ini menggunakan dataset transaksi penjualan produk kopi pada UMKM di Provinsi Lampung dengan total 100 instansi (data transaksi) dan 8 variabel/fitur tanpa missing values. Rincian variabel fitur dalam dataset tersebut meliputi: jenis kopi, grade kopi, metode proses, harga per kg, channel penjualan, status sertifikasi, rating kepuasan dan status keberhasilan sebagai variable target. Dataset ini mencerminkan realitas operasional UMKM dengan merekam berbagai variabel

EXPLORE: ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: http://dx.doi.org/10.36448/jsit.v17i1.5000

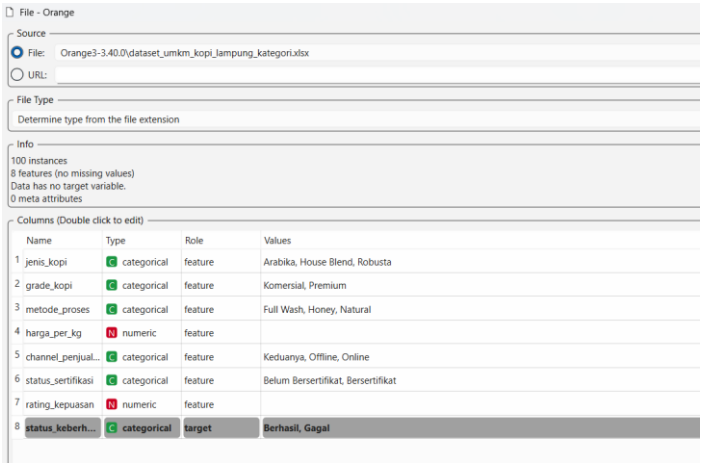
krusial yang mempengaruhi keputusan pembelian dan keberhasilan produk di pasar. Melalui pemrosesan data yang sistematis, dataset ini berfungsi sebagai input utama untuk melatih algoritma agar dapat mengenali pola transaksi yang membedakan produk dengan status penjualan berhasil dan produk yang gagal.



Gambar 3. Dataset

B. Data Preprocessing

Tahapan *data preprocessing* merupakan proses krusial untuk memastikan bahwa data transaksi yang terkumpul memiliki kualitas yang bersih, konsisten, dan siap diolah oleh algoritma *Random Forest*. Mengingat data mentah yang diperoleh dari UMKM kopi sering kali memerlukan penyesuaian format agar dapat dibaca dengan optimal oleh sistem, proses ini dilakukan secara sistematis menggunakan *software* Orange Data Mining. Langkah ini bertujuan untuk memvalidasi struktur data, memastikan penetapan peran variabel yang tepat, serta menjaga integritas informasi sebelum model klasifikasi dibangun

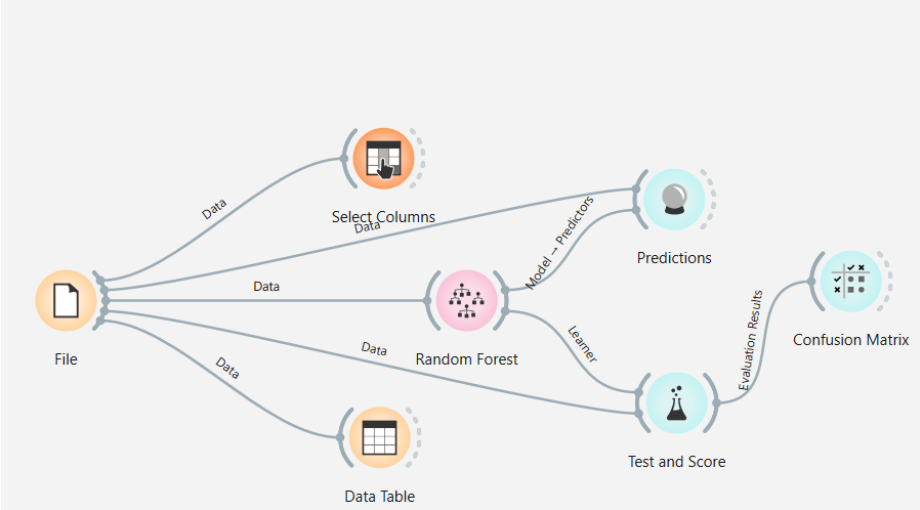


Gambar 4. Data Preprocessing

C. Implementasi Random Forest

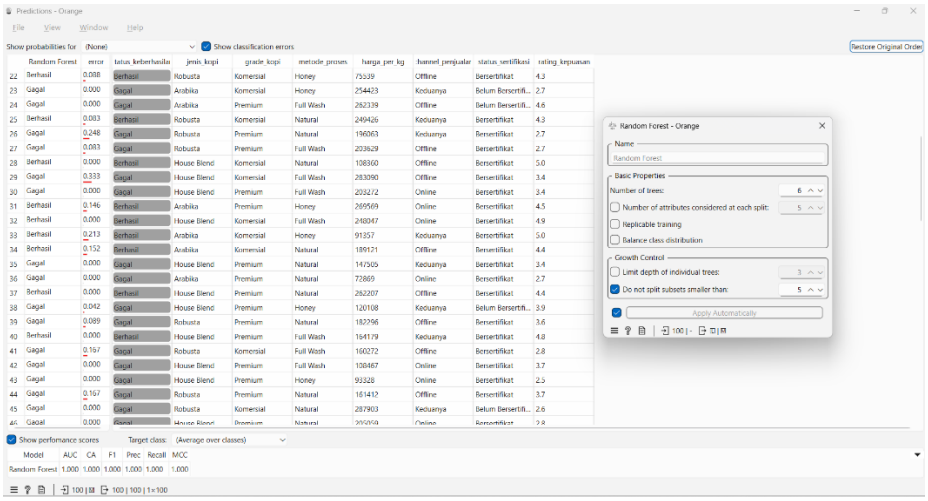
Tahap implementasi model merupakan inti dari penelitian ini, di mana algoritma *Random Forest* diterapkan untuk membangun sistem klasifikasi yang mampu memetakan hubungan kompleks antara variabel penentu dan status keberhasilan penjualan produk kopi. Setelah data dipersiapkan dan divalidasi, proses selanjutnya adalah membangun model klasifikasi yang dapat melakukan generalisasi pola berdasarkan data historis yang tersedia. Dengan menggunakan *software Orange Data Mining*, implementasi dilakukan melalui alur kerja yang terstruktur guna memastikan bahwa model yang dihasilkan tidak hanya akurat secara matematis, tetapi juga dapat diinterpretasikan dengan baik untuk kebutuhan pengambilan keputusan strategis di UMKM kopi Lampung.

EXPLORE: ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: http://dx.doi.org/10.36448/jsit.v17i1.5000



.

Gambar 5. *Data Preprocessing*



Gambar 6. Prediksi *Random Forest*

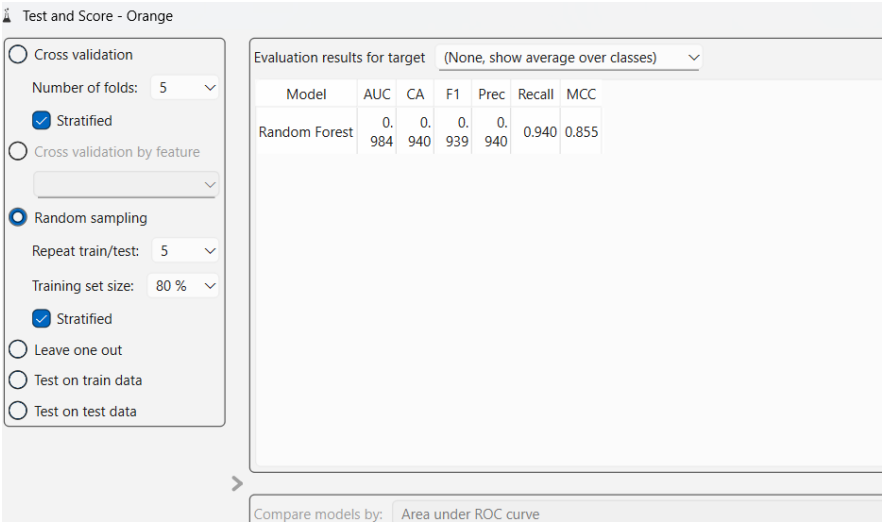
Gambar 6 menampilkan antarmuka hasil prediksi (*Predictions*) dari model *Random Forest* yang dibangun di dalam perangkat lunak *Orange Data Mining*. Kolom-kolom yang disajikan memuat perbandingan antara nilai target aktual (*status_keberhasilan*) dengan hasil prediksi model, dilengkapi dengan nilai probabilitas kesalahan (*error*) untuk setiap instansi data transaksi. Berdasarkan visualisasi tersebut, terlihat bahwa sebagian besar data transaksi berhasil diprediksi secara tepat oleh model (ditandai dengan nilai *error* yang bernilai nol atau sangat kecil di bawah ambang batas toleransi), yang mengindikasikan bahwa algoritma *Random Forest* mampu mengenali pola karakteristik produk kopi (seperti jenis, harga, dan rating) secara konsisten dan akurat sebelum dilanjutkan ke tahap evaluasi metrik pengujian.

D. Evaluasi Model

Evaluasi model dimulai dengan pengujian performa menggunakan data uji (*testing set*) guna memastikan kemampuan generalisasi model terhadap data baru secara objektif.

1) Test and Score

Tahap evaluasi model merupakan langkah kritis dalam penelitian ini untuk mengukur sejauh mana algoritma *Random Forest* yang telah dilatih mampu mengklasifikasikan keberhasilan penjualan produk kopi secara akurat. Evaluasi ini dilakukan menggunakan teknik Random Sampling dengan rasio pelatihan 80% dan pengujian 20% melalui *widget Test and Score* pada *software Orange Data Mining*. Tujuannya adalah untuk memvalidasi performa model menggunakan metrik standar industri machine learning, sehingga tingkat keandalan model dalam memprediksi data yang belum pernah dilihat sebelumnya dapat terukur secara objektif.



Gambar 7. Test and Score

Penjabaran hasil evaluasi berdasarkan *widget Test and Score* menunjukkan performa model *Random Forest* yang sangat solid dalam melakukan klasifikasi. Nilai *Classification Accuracy* (CA) sebesar 0.940 mengindikasikan bahwa model memiliki tingkat ketepatan klasifikasi sebesar 94%, yang menunjukkan bahwa sistem mampu membedakan antara transaksi "Berhasil" dan "Gagal" dengan sangat baik. Selain itu, nilai *Area Under the Curve* (AUC) sebesar 0.984 mencerminkan kemampuan model yang luar biasa dalam memisahkan kedua kelas tersebut. Metrik pendukung lainnya seperti *F1-Score* (0.939), *Precision* (0.940), dan *Recall* (0.940) memberikan bukti konsistensi bahwa model tidak hanya akurat secara keseluruhan, tetapi juga memiliki keseimbangan yang baik dalam meminimalkan kesalahan prediksi (baik *False Positive* maupun *False Negative*). Nilai MCC (*Matthews Correlation Coefficient*) sebesar 0.855 semakin memperkuat keyakinan bahwa model yang dibangun memiliki korelasi yang sangat kuat antara prediksi yang dihasilkan dengan realitas data sebenarnya, menjadikannya model yang layak dan andal untuk diterapkan sebagai instrumen pendukung keputusan strategis bagi UMKM kopi di Lampung.

2) *K-Fold Cross Validation*

Untuk memastikan bahwa performa tinggi yang dihasilkan oleh model *Random Forest* bersifat konsisten dan tidak hanya bergantung pada satu skema pembagian data tertentu yang kebetulan optimal, proses pengujian dalam penelitian ini diperkuat menggunakan teknik *K-Fold Cross Validation*. Pendekatan ini dilakukan dengan membagi keseluruhan dataset ke dalam 5 bagian lipatan (*folds*) yang bersifat *stratified*, di mana proporsi kelas "Berhasil" dan "Gagal" dijaga agar tetap seimbang pada setiap bagiannya. Proses pelatihan (*training*) dan pengujian (*testing*) dilakukan secara berulang sebanyak 5 iterasi secara bergantian di mana 4 bagian lipatan digunakan untuk melatih model dan 1 bagian sisanya digunakan sebagai data uji. Hasil evaluasi dari kelima iterasi tersebut kemudian diagregasi dan dirata-ratakan untuk menguji stabilitas model. Penggunaan *K-Fold Cross Validation* ini membuktikan bahwa metrik performa model teruji sangat stabil, konsisten, dan memiliki kemampuan generalisasi yang baik terhadap data baru di luar data latih.

3) *Confusion Matrix*

Tahapan evaluasi model tidak lengkap tanpa melakukan pembedahan terhadap distribusi kesalahan klasifikasi yang dilakukan oleh model. *Confusion Matrix* digunakan sebagai instrumen analitik untuk memetakan performa model dalam mengenali setiap kelas target secara mendetail. Dengan melihat matriks ini, kita dapat mengetahui secara spesifik berapa banyak transaksi yang berhasil diprediksi dengan tepat dan di mana letak kelemahan model saat melakukan klasifikasi terhadap data "Berhasil" maupun "Gagal". Berdasarkan *Confusion Matrix* yang dihasilkan dari pengujian 100 data transaksi, dapat dijabarkan performa prediksi model *Random Forest* sebagai berikut

EXPLORE: ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: http://dx.doi.org/10.36448/jsit.v17i1.5000

Gambar 8. Test and Score

a) *True Positive* (TP): Terdapat 26 transaksi yang sebenarnya "Berhasil" dan berhasil diprediksi oleh model sebagai "Berhasil".
b) *True Negative* (TN): Terdapat 68 transaksi yang sebenarnya "Gagal" dan berhasil diprediksi oleh model sebagai "Gagal".
c) *False Positive* (FP): Terdapat 2 transaksi yang sebenarnya "Gagal", namun keliru diprediksi oleh model sebagai "Berhasil".
d) *False Negative* (FN): Terdapat 4 transaksi yang sebenarnya "Berhasil", namun keliru diprediksi oleh model sebagai "Gagal".

Secara keseluruhan, model menunjukkan tingkat akurasi yang sangat baik dengan total kesalahan hanya sebesar 6 data (4 *False Negative* dan 2 *False Positive*) dari total 100 data transaksi. Dominasi nilai pada diagonal utama (*True Positive* dan *True Negative*) menunjukkan bahwa algoritma *Random Forest* telah berhasil mengenali pola perilaku penjualan kopi pada UMKM Lampung secara signifikan. Rendahnya angka kesalahan prediksi ini memberikan tingkat kepercayaan yang tinggi bagi pelaku UMKM untuk mengadopsi model ini sebagai alat bantu dalam memprediksi status keberhasilan penjualan produk mereka di masa mendatang.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma *Random Forest* berhasil memberikan solusi yang efektif dan akurat dalam mengklasifikasikan keberhasilan penjualan produk kopi pada UMKM di Lampung. Dengan tingkat akurasi mencapai 94% dan nilai *Area Under the Curve* (AUC) sebesar 0,984, model yang dibangun terbukti memiliki kemampuan generalisasi yang sangat baik dalam mengenali pola transaksi yang kompleks. Penelitian ini memiliki implikasi teoritis yang memperkuat literatur *data science* dan *machine learning* terapan, khususnya dalam membuktikan bahwa algoritma berbasis *ensemble learning* seperti *Random Forest* sangat tangguh dalam menangani data transaksional sektor agrikultur dan hilirisasi produk lokal tanpa rentan terhadap *overfitting*. Selain itu, analisis *feature importance* memberikan kontribusi teoretis dalam memetakan hierarki variabel (seperti dominasi *rating kepuasan* dan *harga per kg*) yang memengaruhi dinamika pasar kopi regional. Dari sisi implikasi praktis, model prediktif ini berhasil mentransformasikan proses pengambilan keputusan pelaku UMKM dari yang semula berbasis intuisi subjektif (*trial and error*) menjadi pendekatan berbasis data (*data-driven decision making*). Hal ini memungkinkan pelaku usaha untuk mengoptimalkan manajemen stok, menetapkan strategi harga yang kompetitif, serta meningkatkan efisiensi operasional guna mendongkrak daya saing di pasar nasional maupun global.

Sehubungan dengan hasil dan keterbatasan penelitian ini, diajukan beberapa saran yang lebih spesifik guna pengembangan model klasifikasi dan penelitian lanjutan ke depan. Pertama, bagi peneliti selanjutnya, disarankan untuk melakukan eksperimen lanjutan menggunakan teknik penyeimbangan data (seperti *SMOTE* jika terdapat potensi ketimpangan kelas pada dataset yang lebih besar) serta melakukan *hyperparameter tuning* yang lebih mendalam (seperti optimasi *Grid Search* atau *Random Search* pada parameter n_estimators, max_depth, dan *criterion*) guna menguji apakah performa akurasi dapat ditingkatkan melampaui angka 94%. Kedua, cakupan variabel penelitian dapat diperluas dengan menambahkan faktor-faktor eksternal yang memengaruhi daya beli konsumen, seperti tren musiman, data

demografi pembeli, efektivitas kampanye *digital marketing*, serta analisis sentimen ulasan pelanggan secara *real-time*. Terakhir, model klasifikasi *Random Forest* yang telah teruji ini direkomendasikan untuk diintegrasikan ke dalam sebuah ekosistem perangkat lunak nyata seperti aplikasi berbasis *web* atau *mobile* yang ramah pengguna (*user-friendly*) sehingga para pelaku UMKM kopi di Lampung dapat mengakses dan memanfaatkan teknologi prediktif ini secara mandiri dalam kegiatan operasional harian mereka tanpa memerlukan keahlian teknis yang mendalam.

5. Daftar Pustaka

[1] C. H. Sumerli, M. M. Syuhada, and E. Karundeng, "Implementasi Model Bisinis Inovatif dan Penggunaan BigData dalam Meningkatkan Efisiensi Operasional serta Keberlanjutan UMKM Nasional," *J. Econ. Stud.*, vol. 1, no. 2, pp. 93–106, 2025.

[2] A. Maharani, "Penerapan Big data dalam Perencanaan Strategis dan Pengambilan Keputusan Bisnis," *JMEB J. Manaj. Ekon. Bisnis*, vol. 3, no. 1, pp. 15–23, 2025, [Online]. Available: https://journal.sabajayapublisher.com/index.php/jmeb

[3] E. M. A. Kadir, "Transformasi Digital Umkm Melalui Pemanfaatan Artificial Intelligence Dalam Meningkatkan Daya Saing Bisnis Di Era Ekonomi Digital," *J. Inov. Pembelajaran dan Teknol. Mod.*, vol. 9, no. 4, pp. 37–58, 2025.

[4] Dio Samudra and A. A. Saputra, "Perkembangan Ekonomi Dan Pembangunan Manusia Di Provinsi Lampung," *jurnal.balitbangda.lampungprov.go.id*, vol. 14, no. 1, pp. 15–24, 2026, [Online]. Available: http://www.undp.org.lb/programme/governance/advocacy/nhdr/nhdr98/chptr1.pdf

[5] A. Niravita, "Peningkataan Kapasitas dan Daya Saing UMKM Kopi Di Kabupaten Temanggung Melalui Legalitas Usaha," *J. Pengabdi. Kpd. Masy. Nusant.*, vol. 5, no. 4, pp. 4653–4664, 2024.

[6] M. D. Ubaidilah, S. D. Mutia, and E. Daniati, "Klasifikasi Pola Penjualan Berdasarkan Waktu dengan Metode Naive Bayes untuk Meningkatkan Efisiensi Penjualan Toko Lia Batik Collection Kediri," *INOTEK*, vol. 9, pp. 1760–1771, 2024.

[7] S. Dewi, H. A. Al Kautsar, and D. Y. Utami, "Prediksi Keberhasilan Pemasaran Layanan Jasa Perbankan Mengunnakan Algoritma Logistic Regreesion," *Comput. Sci.*, vol. 3, no. 2, pp. 118–125, 2023, doi: 10.31294/coscience.v3i2.1931.

[8] R. Setiawan, B. Priyatna, E. Novalia, and B. Huda, "Klasifikasi Tingkat Penjualan Produk pada Toko Jati Karebet Menggunakan Algoritma Naïve Bayes," *J. Fasilkom*, vol. 15, no. 2, pp. 297–303, 2025.

[9] M. R. Romahdoni and M. A. K. Wardana, "Penerapan Business Intelligence Terhadap Strategi Pengembangan Produk Unggul Pada UMKM Ecoprint Menggunakan Algoritma Apriori," *J. Inform.*, vol. 24, no. 2, pp. 94–107, 2024.

[10] R. A. Maatuil, "Penerapan Model Analisis Risiko Kredit dalam Pengambilan Keputusan Pembiayaan pada UMKM," *Indones. Econ. J.*, vol. 1, no. 2, pp. 635–664, 2025.

[11] M. R. Romahdoni and M. A. K. Wardana, "Aplikasi Business Intelligence Berbasis Artificial Intelligence untuk Prediksi Keberhasilan Penjualan Produk dan Keberlanjutan UMKM," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 15, no. 2, pp. 202–212, 2025.

[12] A. P. A. Prayogi, A. I. Shofyana, and D. Putriani, "Prediksi Credit Card Approval Menggunakan Algoritma Random Forest," *J. Ilmu Komput. dan Multimed.*, vol. 2, no. 1, pp. 21–25, 2025.

[13] M. U. Sattar, V. Dattana, R. Hasan, S. Mahmood, H. W. Khan, and S. Hussain, "Enhancing Supply Chain Management: A Comparative Study of Machine Learning Techniques with Cost–Accuracy and ESG-Based Evaluation for Forecasting and Risk Mitigation," *Sustain.*, vol. 17, no. 13, pp. 1–45, 2025, doi: 10.3390/su17135772.

[14] Y. Cao, L. Hao, Z. Zhang, and H. Zhang, "Path Exploration of Artificial Intelligence-Driven Green Supply Chain Management in Manufacturing Enterprises: A Study Based on Random Forest and Dynamic QCA Under the TOE Framework," *Systems*, vol. 13, no. 12, 2025, doi: 10.3390/systems13121120.

[15] Hidayat and A. Sunyoto, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 7, no. September, pp. 31–40, 2023.

[16] S. Helmiyah, "Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi, Presisi, Recall, dan F1-Score pada Dataset Kacang Kering," *J. Ilmu Komput. Dan Teknol.*, vol. 6, no. 3, pp. 152–159, 2025, [Online]. Available: https://doi.org/10.35960/ikomti.v6i3.2031%0Ahttp://creativecommons.org/licences/by/4.0/%0Ahttp://ejournal.uhb.ac.id/index.php/IKOMTI

[17] M. Hussein Umar, R. Daud Antony Pangaribuan, R. Primadana, I. Budiawan, and R. Pakpahan, "Penerapan Algoritma Random Forest Untuk Prediksi Tingkat Stres Dari Aktivitas Media Sosial," *J. Media Inform. [JUMIN]*, vol. 6, no. 6, pp. 3113–3112, 2025.