



PERBANDINGAN ALGORITMA RANDOM FOREST DAN DECISION TREE UNTUK KLASIFIKASI PERTUMBUHAN UMKM KULINER DI KOTA MANADO DENGAN MENGGUNAKAN PENDEKATAN EXPLAINABLE AI

Keren Aprilia Mumbunan^{1*}, Arie S Lumenta², Yaulie Rindengan³

^{1,2,3} Teknik Informatika, Universitas Sam Ratulangi Manado, Indonesia

E-mail: kerenmumbunan026@student.unsrat.ac.id¹, al@unsrat.ac.id²,
rindengan@unsrat.ac.id³

ABSTRACT

Culinary MSMEs play a significant role in Manado City's creative economy, yet their growth still faces various internal and external obstacles that are difficult to identify objectively. This study aims to build a classification model for the growth level of culinary MSMEs in Manado City using Random Forest and Decision Tree algorithms, compare their performance based on accuracy, precision, recall, and F1-score, and interpret the most influential factors behind the classification results through an Explainable AI (XAI) approach. Primary data were collected through a questionnaire distributed to 101 culinary MSME owners, consisting of 16 predictor features and one target variable representing growth level (Low, Medium, High) derived from revenue change. The data were processed through encoding, missing-value imputation, stratified 80:20 train-test splitting. Both algorithms were optimized through Grid Search hyperparameter tuning and evaluated using Repeated Stratified K-Fold Cross-Validation. Test results show that Random Forest achieved an accuracy of 61.90% with a weighted F1-score of 0.5929, while Decision Tree achieved an accuracy of 52.38% with a weighted F1-score of 0.4812, establishing Random Forest as the best-performing model. SHAP and LIME analyses on both models consistently identified financial record-keeping, education level, business duration, capital source, and social media promotion activity as the factors most influential to MSME growth. The best-performing model was subsequently implemented as a Streamlit-based web application deployed on Hugging Face Spaces for direct use by MSME owners and policymakers.

Keywords: *culinary MSME, Random Forest, Decision Tree, Explainable AI, SHAP, business growth classification*

Article History

Received: Juli 2026

Reviewed: Juli 2026

Published: Juli 2026

Plagiarism Checker No
237.u873247.fh8374

Copyright : Author

Publish by : Kohesi



This work is licensed
under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

**ABSTRAK**

UMKM kuliner merupakan salah satu sektor ekonomi kreatif yang berperan penting bagi perekonomian Kota Manado, namun pertumbuhannya masih menghadapi berbagai hambatan internal dan eksternal yang sulit diidentifikasi secara objektif. Penelitian ini bertujuan membangun model klasifikasi tingkat pertumbuhan UMKM kuliner di Kota Manado menggunakan algoritma Random Forest dan Decision Tree, membandingkan performa kedua algoritma berdasarkan metrik accuracy, precision, recall, dan F1-score, serta menginterpretasikan faktor-faktor paling berpengaruh terhadap hasil klasifikasi melalui pendekatan Explainable Artificial Intelligence (XAI). Data primer diperoleh dari kuesioner yang disebarakan kepada 101 pelaku UMKM kuliner, terdiri atas 16 fitur predictor dan satu variabel target tingkat pertumbuhan (Rendah, Sedang, Tinggi) yang dibentuk berdasarkan perubahan omzet. Data diproses melalui encoding, imputasi nilai hilang, pembagian data latih-uji 80:20 secara stratified. Kedua algoritma dioptimalkan melalui hyperparameter tuning dengan Grid Search dan dievaluasi menggunakan Repeated Stratified K-Fold Cross-Validation. Hasil pengujian menunjukkan Random Forest memperoleh akurasi 61,90% dengan F1-score weighted 0,5929, sedangkan Decision Tree memperoleh akurasi 52,38% dengan F1-score weighted 0,4812, sehingga Random Forest ditetapkan sebagai model terbaik. Analisis Shapley Additive Explanations (SHAP) dan Local Interpretable Model-agnostic Explanations (LIME) pada kedua model mengidentifikasi pencatatan keuangan, pendidikan, lama usaha, sumber modal, dan keaktifan promosi media sosial sebagai faktor yang secara konsisten memengaruhi tingkat pertumbuhan UMKM. Model terbaik selanjutnya diimplementasikan sebagai aplikasi web berbasis Streamlit yang di-deploy melalui Hugging Face Spaces agar dapat digunakan langsung oleh pelaku UMKM maupun pemangku kebijakan.

Kata kunci: UMKM kuliner, Random Forest, Decision Tree, Explainable Artificial Intelligence (XAI), SHAP, klasifikasi pertumbuhan usaha

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan kegiatan usaha produktif yang dimiliki oleh perorangan maupun badan usaha perorangan yang memenuhi kriteria tertentu berdasarkan aset, omzet, dan jumlah tenaga kerja sebagaimana diatur dalam Undang-Undang Republik Indonesia Nomor 20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah. UMKM memiliki peran strategis



dalam perekonomian Indonesia sebagai tulang punggung pertumbuhan ekonomi nasional [1] dengan kontribusi signifikan terhadap Produk Domestik Bruto (PDB), penyerapan tenaga kerja, dan pemerataan pendapatan masyarakat [2]. Data Kementerian Koperasi dan Usaha Kecil dan Menengah Republik Indonesia mencatat jumlah UMKM mencapai 64,2 juta unit atau sekitar 99,99% dari total pelaku usaha nasional, menyerap sekitar 117 juta tenaga kerja, serta berkontribusi sebesar 61,1% terhadap PDB nasional [3].

Di antara berbagai sektor UMKM, sektor kuliner menempati posisi yang strategis karena merupakan kontributor terbesar terhadap PDB ekonomi kreatif Indonesia. Kemampuan UMKM kuliner beradaptasi terhadap perubahan tren pasar dan melakukan inovasi produk menjadi faktor penting yang mendukung peningkatan kinerja usaha [4]. Di Kota Manado, perkembangan UMKM kuliner didukung oleh keberagaman produk lokal, aktivitas perdagangan dan pariwisata, serta meningkatnya transformasi digital yang dilakukan pelaku usaha [5], sehingga menjadikan kota ini lokasi yang relevan untuk mengkaji pola pertumbuhan UMKM kuliner secara lebih mendalam.

Meskipun memiliki potensi besar, pertumbuhan UMKM kuliner masih menghadapi berbagai tantangan, di antaranya keterbatasan akses terhadap pembiayaan formal yang menghambat peningkatan kapasitas usaha, adopsi teknologi, dan perluasan pasar [6], serta rendahnya literasi digital dan keterbatasan kemampuan manajerial yang menghambat transformasi digital dan peningkatan kinerja usaha [7]. Kondisi ini menunjukkan perlunya pendekatan berbasis data untuk memahami pola pertumbuhan UMKM kuliner serta mengidentifikasi faktor-faktor yang memengaruhinya sebagai dasar penyusunan strategi pengembangan yang lebih tepat sasaran.

Pendekatan analisis konvensional memiliki keterbatasan dalam menangkap hubungan kompleks antarvariabel yang dipengaruhi berbagai faktor internal dan eksternal yang saling berinteraksi, sehingga penerapan algoritma machine learning semakin banyak digunakan karena mampu menghasilkan model klasifikasi dengan akurasi lebih baik dibandingkan pendekatan konvensional pada berbagai permasalahan bisnis dan ekonomi [8]. Decision Tree dan Random Forest merupakan dua algoritma klasifikasi yang banyak diterapkan pada data tabular. Decision Tree membangun model melalui struktur pohon keputusan yang mudah dipahami, sedangkan Random Forest mengembangkan pendekatan ensemble dengan mengombinasikan banyak pohon keputusan untuk meningkatkan stabilitas model, mengurangi risiko overfitting, dan meningkatkan akurasi klasifikasi [9]. Pemilihan algoritma machine learning terbukti memberikan pengaruh terhadap kualitas hasil klasifikasi UMKM [10], sehingga evaluasi terhadap beberapa algoritma perlu dilakukan untuk memperoleh model yang paling sesuai dengan karakteristik data.

Meskipun Random Forest umumnya memberikan performa lebih baik, algoritma ini tergolong *black box model* karena kompleksitas ratusan hingga ribuan pohon keputusan di dalamnya membuat proses pengambilan keputusan sulit ditelusuri [11], sedangkan Decision Tree yang lebih mudah dipahami cenderung memiliki keterbatasan akurasi pada data yang kompleks atau memiliki variansi tinggi. Pemangku kepentingan UMKM tidak hanya membutuhkan hasil klasifikasi, tetapi juga pemahaman mengenai faktor-faktor yang paling berpengaruh terhadap pertumbuhan usaha, sehingga pendekatan Explainable Artificial Intelligence (XAI) menjadi penting untuk meningkatkan transparansi dan interpretabilitas model machine learning yang kompleks [12]. Teknik SHAP (SHapley Additive exPlanations) mampu menjelaskan kontribusi tiap



variabel terhadap hasil prediksi secara konsisten pada tingkat lokal maupun global [13], sementara LIME (Local Interpretable Model-agnostic Explanations) membangun model penjelas sederhana di sekitar satu prediksi sehingga menghasilkan interpretasi lokal yang mudah dipahami terhadap keputusan model black box.

Penelitian terkait klasifikasi kinerja dan pertumbuhan usaha berbasis machine learning hingga saat ini umumnya masih berfokus pada perusahaan berskala besar, seperti pengelompokan kinerja keuangan PT Pegadaian menggunakan K-Means dengan Random Forest sebagai metode validasi klaster [14], atau menggunakan data agregat berskala nasional. Penelitian yang secara khusus mengkaji UMKM kuliner di tingkat kota dengan membandingkan algoritma Random Forest dan Decision Tree sekaligus mengintegrasikan pendekatan Explainable AI masih sangat terbatas dilakukan di Indonesia, khususnya di Kota Manado, sehingga membuka peluang penelitian untuk mengisi kesenjangan tersebut.

Berdasarkan uraian tersebut, penelitian ini bertujuan membangun model klasifikasi tingkat pertumbuhan UMKM kuliner di Kota Manado menggunakan algoritma Random Forest dan Decision Tree, membandingkan performa kedua algoritma berdasarkan metrik akurasi, precision, recall, dan F1-score, serta menginterpretasikan faktor-faktor yang paling berpengaruh terhadap tingkat pertumbuhan usaha melalui pendekatan Explainable AI berbasis SHAP dan LIME. Data yang digunakan merupakan data primer yang diperoleh melalui kuesioner terstruktur yang disebarakan kepada pelaku UMKM kuliner di Kota Manado. Penelitian ini diharapkan memberikan solusi berbasis teknologi yang mendukung pengembangan UMKM kuliner yang berkelanjutan dan berdaya saing, sekaligus memberikan gambaran yang transparan dan mudah dipahami bagi pemangku kepentingan mengenai faktor-faktor yang menentukan tingkat pertumbuhan usaha kuliner di Kota Manado.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain klasifikasi berbasis machine learning untuk membandingkan performa algoritma Random Forest dan Decision Tree dalam mengklasifikasikan tingkat pertumbuhan UMKM kuliner, yang selanjutnya diinterpretasikan menggunakan pendekatan Explainable AI (XAI). Penelitian dilaksanakan pada UMKM kuliner di Kota Manado, Provinsi Sulawesi Utara, selama enam bulan (Januari-Juni 2026), mencakup tahapan studi literatur, pengumpulan dan pengolahan data, perancangan serta pelatihan model, hingga penyusunan laporan akhir.

Data yang digunakan merupakan data primer yang dikumpulkan melalui kuesioner terstruktur berbasis Google Forms yang disebarakan kepada pelaku UMKM kuliner di Kota Manado, menghasilkan 101 sampel data. Setiap sampel terdiri atas 16 fitur prediktor yang dikelompokkan ke dalam tiga aspek, yaitu karakteristik usaha dan pemilik (pendidikan, lama usaha, jumlah karyawan, dan modal awal), aspek produksi dan pemasaran (lokasi usaha, media penjualan, hari berjualan, omzet bulanan, dan transaksi per hari), serta aspek dukungan, digitalisasi, dan lingkungan usaha (pencatatan keuangan, sumber modal, penggunaan teknologi, kepemilikan izin usaha, intensitas promosi media sosial, kepemilikan target penjualan, dan keikutsertaan pelatihan). Variabel target tingkat_pertumbuhan dibentuk berdasarkan perubahan omzet



dibandingkan periode sebelumnya dan dikategorikan ke dalam tiga kelas, yaitu Rendah, Sedang, dan Tinggi.

Sebelum pemodelan dilakukan, dataset dianalisis secara eksploratif untuk memahami distribusi tiap variabel, korelasi antarfitur, serta mengidentifikasi nilai hilang dan pencilaan sebagai acuan penentuan strategi preprocessing. Tahap preprocessing meliputi pembersihan data dan penyeragaman nama kolom, pembentukan label target dari variabel perubahan omzet, transformasi fitur kategorikal ke bentuk numerik melalui ordinal encoding, nominal encoding, dan binary encoding sesuai karakteristik masing-masing variabel, serta penanganan nilai hilang pada fitur numerik melalui imputasi median. Dataset kemudian dibagi menjadi data latih dan data uji menggunakan fungsi `train_test_split()` pada library `scikit-learn` dengan parameter `stratify=y`, sehingga proporsi ketiga kelas target tetap terjaga pada kedua subset data. Pembagian ini menghasilkan 80 sampel data latih dan 21 sampel data uji.

Model klasifikasi dibangun menggunakan dua algoritma. Decision Tree membangun struktur pohon keputusan yang bermula dari root node berisi seluruh data, kemudian bercabang ke internal node berdasarkan pengujian fitur yang paling baik memisahkan kelas, hingga berhenti pada leaf node yang berisi hasil prediksi. Random Forest mengembangkan pendekatan tersebut melalui skema ensemble dengan membangun sejumlah besar pohon keputusan menggunakan bootstrap sampling pada data serta pemilihan fitur secara acak pada setiap pemisahan node, kemudian menggabungkan hasil prediksi seluruh pohon melalui majority voting untuk menghasilkan keputusan akhir yang lebih stabil dan tahan terhadap overfitting.

Pencarian kombinasi parameter optimal (hyperparameter tuning) pada kedua algoritma dilakukan menggunakan GridSearchCV yang diterapkan hanya pada data latih untuk menghindari kebocoran data (data leakage) ke data uji. Ruang pencarian parameter Decision Tree meliputi `criterion`, `max_depth`, `min_samples_split`, dan `min_samples_leaf`, sedangkan ruang pencarian Random Forest mencakup parameter yang sama ditambah `n_estimators` dan `max_features`. Mengingat ukuran dataset yang relatif kecil, estimasi performa pada tahap tuning menggunakan skema Repeated Stratified K-Fold Cross-Validation dengan 5 fold yang diulang sebanyak 3 repetisi, untuk memperoleh estimasi performa yang lebih stabil dan menjaga proporsi kelas pada setiap fold pelatihan maupun validasi.

Model dengan kombinasi parameter terbaik hasil tuning kemudian dilatih ulang (refit) menggunakan seluruh data latih, lalu diuji pada 21 sampel data uji yang tidak dilibatkan dalam proses pelatihan maupun tuning. Performa model dievaluasi menggunakan metrik `accuracy`, `precision`, `recall`, dan `F1-score` per kelas yang dihitung dengan pendekatan macro average dan weighted average untuk mengakomodasi ketidakseimbangan jumlah sampel antarkelas, confusion matrix untuk memvisualisasikan pola prediksi terhadap ketiga kelas target, serta ROC curve dan AUC dengan pendekatan one-vs-rest untuk mengukur kemampuan diskriminasi model. Khusus pada Random Forest, dilakukan pula analisis Out-of-Bag (OOB) Score terhadap variasi jumlah pohon (`n_estimators`) sebagai validasi internal tambahan.

Model dengan performa terbaik selanjutnya diinterpretasikan menggunakan pendekatan Explainable AI, yaitu SHAP (SHapley Additive exPlanations) melalui summary plot, force plot, dan dependence plot untuk menjelaskan kontribusi fitur secara global maupun lokal, serta LIME



(Local Interpretable Model-agnostic Explanations) untuk menghasilkan penjelasan lokal berbasis sampel individual. Seluruh tahapan preprocessing, pemodelan, evaluasi, dan interpretasi XAI diimplementasikan menggunakan bahasa pemrograman Python melalui Jupyter Notebook dan Visual Studio Code. Model dengan performa terbaik selanjutnya diimplementasikan ke dalam aplikasi web interaktif berbasis Streamlit yang di-deploy melalui Hugging Face Spaces, sehingga dapat diakses secara publik oleh pelaku UMKM maupun pemangku kebijakan tanpa memerlukan instalasi lokal.

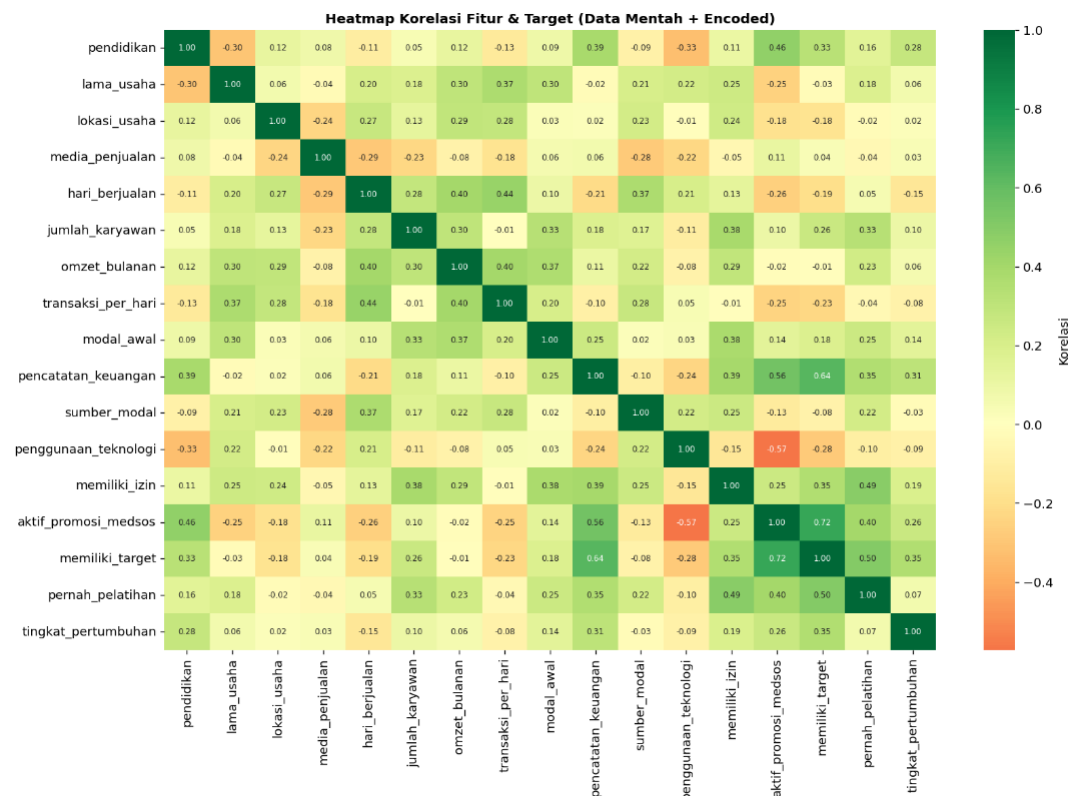
3. HASIL DAN PEMBAHASAN

Deskripsi Dataset dan Eksplorasi Data

Dataset akhir yang digunakan dalam pemodelan terdiri atas 101 sampel UMKM kuliner di Kota Manado tanpa nilai hilang pada seluruh kolom, dengan distribusi kelas target Rendah sebanyak 21 sampel, Sedang 37 sampel, dan Tinggi 43 sampel. Hasil eksplorasi data menunjukkan mayoritas responden berpendidikan SMA/SMK (60 responden), telah menjalankan usaha lebih dari satu tahun (58 dari 101 responden), memiliki omzet bulanan pada rentang Rp1.000.000-Rp5.000.000 (73 responden), berjualan pada lokasi semipermanen seperti stand/booth/gerobak (50 responden), serta masih mengandalkan penjualan offline (72 responden). Pada fitur berskala Likert, intensitas promosi media sosial cenderung tinggi, sedangkan keikutsertaan pelatihan usaha sangat rendah, dengan 49 dari 101 responden belum pernah mengikuti pelatihan sama sekali. Profil ini menggambarkan UMKM kuliner skala kecil dengan operasional yang cukup intensif namun adopsi digital dan akses pelatihan yang masih terbatas.

Analisis Korelasi Antarfitur

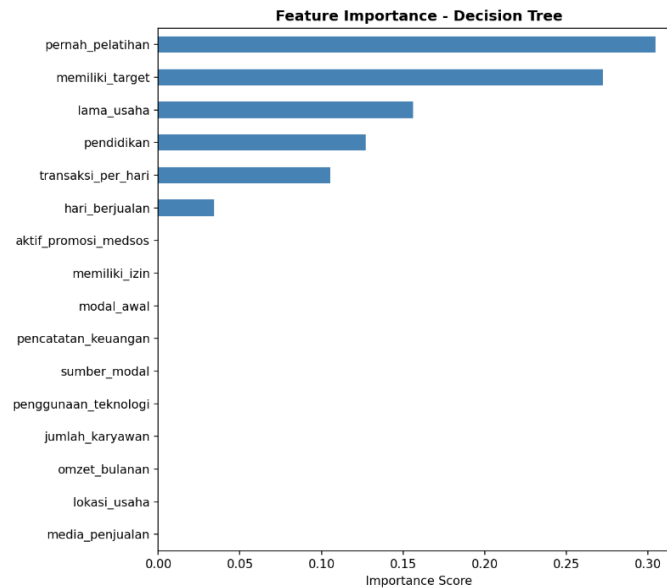
Hasil analisis korelasi menunjukkan bahwa seluruh fitur prediktor memiliki korelasi linear yang tergolong lemah hingga sedang terhadap variabel target, dengan rentang nilai -0,15 hingga 0,31, sebagaimana ditunjukkan pada Gambar 1. Fitur pencatatan_keuangan ($r = 0,31$) dan pendidikan ($r = 0,28$) menunjukkan korelasi positif tertinggi terhadap variabel target, sedangkan fitur lain menunjukkan korelasi yang jauh lebih lemah. Pada matriks korelasi tersebut juga terlihat beberapa pasangan antarfitur prediktor dengan korelasi relatif lebih tinggi, seperti hari_berjualan dengan transaksi_per_hari ($r = 0,44$), namun hal ini tidak menjadi masalah signifikan karena Decision Tree dan Random Forest tidak mengasumsikan independensi antarfitur. Rendahnya nilai korelasi linear pada mayoritas fitur mengindikasikan bahwa hubungan antara karakteristik UMKM dan tingkat pertumbuhan usaha bersifat multifaktorial dan tidak linear, sehingga memperkuat relevansi penggunaan algoritma berbasis pohon keputusan seperti Decision Tree dan Random Forest dibandingkan model linear konvensional.



Gambar 1. Heatmap Korelasi Antarfitur dan Variabel Target

Decision Tree

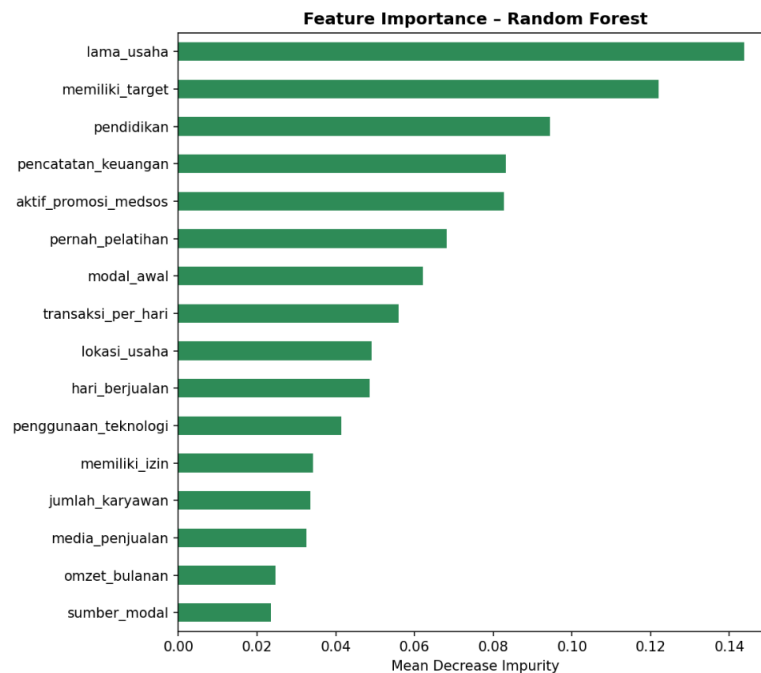
Hyperparameter tuning menggunakan GridSearchCV memilih kombinasi terbaik `criterion='entropy'`, `max_depth=5`, `min_samples_split=2`, dan `min_samples_leaf=4`, dengan akurasi rata-rata cross-validation sebesar 40,00%. Pada data uji, model Decision Tree memperoleh akurasi sebesar 52,38% dengan rata-rata cross-validation 5-fold sebesar 0,4125 ($\pm 0,1205$). Secara per kelas, model sama sekali tidak mampu mengenali kelas Rendah (precision, recall, dan F1-score seluruhnya 0,00) karena hanya terdapat 4 dari 21 sampel kelas tersebut pada data uji, sedangkan kelas Sedang dan Tinggi memperoleh F1-score masing-masing 0,59 dan 0,60. Nilai F1-score rata-rata makro sebesar 0,40 dan rata-rata tertimbang sebesar 0,48. Kegagalan model dalam mengenali kelas Rendah bukan disebabkan kesalahan pemodelan, melainkan konsekuensi dari kecilnya jumlah data uji pada kelas tersebut, sebagai akibat langsung dari ukuran dataset yang terbatas. Berdasarkan analisis feature importance sebagaimana ditunjukkan pada Gambar 2, model Decision Tree sangat bergantung pada sejumlah kecil fitur dominan, yaitu `pernah_pelatihan` (0,30), `memiliki_target` (0,27), `lama_usaha` (0,16), `pendidikan` (0,13), dan `transaksi_per_hari` (0,10), sementara sebagian besar fitur lain memiliki kontribusi mendekati nol.



Gambar 2. Feature Importance Model Decision Tree

Random Forest

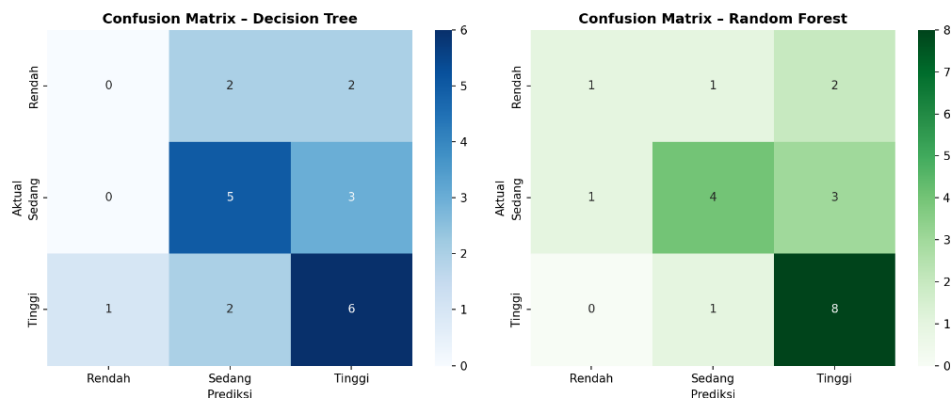
Pada Random Forest, ruang pencarian parameter yang lebih luas menghasilkan kombinasi terbaik $n\_estimators=50$, $max\_depth=None$, $max\_features='log2'$, dan $min\_samples\_leaf=2$, dengan akurasi cross-validation sebesar 46,67%, sedikit lebih tinggi dibandingkan Decision Tree. Pada data uji, model Random Forest memperoleh akurasi sebesar 61,90% dengan rata-rata cross-validation 0,4462 ($\pm 0,1008$), menunjukkan variasi antar-lipatan yang lebih kecil sehingga performanya cenderung lebih stabil. Secara per kelas, kelas Rendah memperoleh precision 0,50, recall 0,25, dan F1-score 0,33, yang menunjukkan perbaikan signifikan dibandingkan Decision Tree yang gagal total mengenali kelas ini. Kelas Sedang memperoleh F1-score 0,57, sedangkan kelas Tinggi menunjukkan performa terbaik dengan F1-score 0,73. Nilai F1-score rata-rata makro sebesar 0,54 dan rata-rata tertimbang sebesar 0,59. Sebagaimana ditunjukkan pada Gambar 3, Random Forest mendistribusikan kontribusi ke lebih banyak fitur dibandingkan Decision Tree, dengan lama_usaha (0,144), memiliki_target (0,122), pendidikan (0,094), pencatatan_keuangan dan aktif_promosi_medsos (masing-masing sekitar 0,083), serta pernah_pelatihan (0,068) sebagai kontributor utama, sementara seluruh 16 fitur tetap memberikan kontribusi yang lebih besar dari nol. Analisis Out-of-Bag Score terhadap variasi jumlah pohon juga menunjukkan peningkatan tajam dari sekitar 0,33 pada 10-30 pohon menjadi sekitar 0,425 pada 40 pohon, kemudian berfluktuasi stabil pada kisaran 0,40-0,45 hingga mencapai puncak sekitar 0,45 pada 180 pohon, yang mengindikasikan bahwa penambahan jumlah pohon lebih lanjut tidak lagi memberikan peningkatan performa yang berarti.



Gambar 3. Feature Importance Model Random Forest

Perbandingan Performa Model

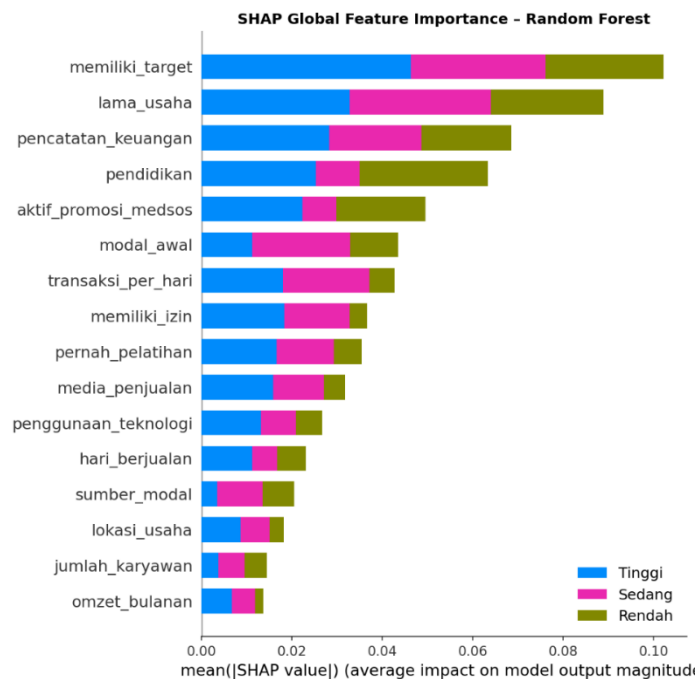
Random Forest secara konsisten mengungguli Decision Tree pada seluruh metrik evaluasi berbasis rata-rata tertimbang: akurasi 61,90% berbanding 52,38%, precision 0,6129 berbanding 0,4454, recall 0,6190 berbanding 0,5238, dan F1-score 0,5929 berbanding 0,4812. Selisih paling mencolok terlihat pada metrik precision, yang mengindikasikan bahwa Random Forest jauh lebih tepat dalam memastikan sampel yang diprediksi pada suatu kelas memang benar berasal dari kelas tersebut. Perbandingan pola kesalahan klasifikasi kedua model pada data uji ditunjukkan pada Gambar 4. Model Decision Tree sama sekali tidak berhasil memprediksi data kelas Rendah dengan benar, sedangkan Random Forest berhasil mengenali satu dari empat data kelas Rendah, serta unggul pada kelas Tinggi dengan delapan dari sembilan data diprediksi benar. Secara keseluruhan, dari 21 data uji, Decision Tree menghasilkan 11 prediksi benar sedangkan Random Forest menghasilkan 13 prediksi benar. Namun demikian, hasil analisis ROC-AUC dengan pendekatan one-vs-rest menunjukkan pola yang lebih bernuansa: Decision Tree memperoleh AUC yang lebih tinggi pada kelas Rendah (0,69) dan Sedang (0,58) dibandingkan Random Forest (0,59 dan 0,61), sedangkan Random Forest unggul pada kelas Tinggi (0,72 berbanding 0,67). Hal ini menunjukkan bahwa meskipun keputusan akhir Decision Tree pada kelas Rendah keliru, model tetap mampu memberikan peringkat probabilitas yang relatif informatif pada seluruh ambang batas, sehingga evaluasi model perlu mempertimbangkan lebih dari satu metrik



Gambar 4. Perbandingan Confusion Matrix Decision Tree dan Random Forest

Interpretasi SHAP

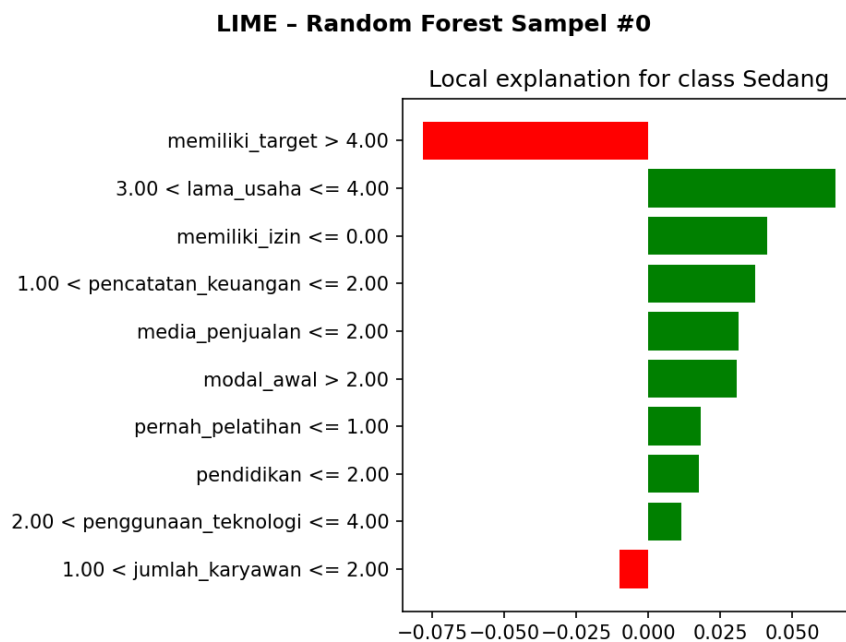
Analisis SHAP global pada Random Forest, sebagaimana ditunjukkan pada Gambar 5, mengonfirmasi memiliki_target, lama_usaha, pencatatan_keuangan, pendidikan, dan aktif_promosi_medsos sebagai fitur yang paling berpengaruh terhadap hasil prediksi, sejalan dengan hasil feature importance bawaan model. Pola pada setiap kelas menunjukkan bahwa usaha yang memiliki target penjualan jelas, rutin melakukan pencatatan keuangan, dan aktif berpromosi di media sosial cenderung didorong ke kelas Tinggi, sedangkan usaha tanpa target dan minim pencatatan keuangan cenderung didorong ke kelas Rendah. Pada Decision Tree, SHAP juga mengonfirmasi memiliki_target, lama_usaha, pendidikan, dan pernah_pelatihan sebagai fitur dominan, namun dengan pola yang lebih terkonsentrasi pada segelintir fitur, menegaskan karakteristik Decision Tree yang membentuk struktur percabangan hanya berdasarkan sebagian kecil fitur paling informatif.



Gambar 5. SHAP Global Feature Importance pada Model Random Forest

Interpretasi Lokal SHAP dan LIME

Analisis lokal melalui SHAP waterfall plot dan LIME pada beberapa sampel individual menunjukkan pola yang konsisten dengan temuan global, di mana kombinasi kepemilikan target usaha dan lama usaha secara berulang muncul sebagai faktor paling berpengaruh terhadap prediksi pada tingkat sampel, baik pada Random Forest maupun Decision Tree. LIME memperkuat temuan tersebut melalui penjelasan berbasis rentang nilai fitur yang lebih sederhana dan mudah dipahami, sebagaimana dicontohkan pada Gambar 6 untuk salah satu sampel data yang diprediksi ke kelas Sedang oleh model Random Forest. Pada contoh tersebut, kondisi memiliki_target > 4,00 memberikan kontribusi negatif paling besar terhadap prediksi kelas Sedang, sedangkan kondisi lama_usaha pada rentang 3,00-4,00 tahun menjadi faktor pendukung utama, yang menegaskan bahwa kombinasi kepemilikan target penjualan dan lama usaha pada rentang tertentu secara konsisten muncul sebagai kondisi paling berpengaruh pada kedua model. Konsistensi antara hasil SHAP dan LIME pada tingkat global maupun lokal memperkuat validitas interpretasi bahwa kepemilikan target usaha, lama usaha, pencatatan keuangan, dan pendidikan merupakan faktor-faktor kunci yang menentukan tingkat pertumbuhan UMKM kuliner di Kota Manado.



Gambar 6. LIME - Random Forest Sampel #0 (Kelas: Sedang)

Pembahasan

Keunggulan Random Forest dibandingkan Decision Tree pada penelitian ini sejalan dengan karakteristik teoretis kedua algoritma, di mana pendekatan ensemble pada Random Forest terbukti lebih stabil dan lebih tahan terhadap overfitting dibandingkan pohon keputusan tunggal, khususnya pada dataset berukuran kecil dengan distribusi kelas yang tidak seimbang. Temuan bahwa pencatatan keuangan, pendidikan, lama usaha, dan keaktifan promosi media sosial menjadi faktor yang paling berpengaruh terhadap pertumbuhan UMKM kuliner memperkuat hasil penelitian terdahulu yang juga mengidentifikasi variabel keuangan dan literasi sebagai faktor determinan kinerja UMKM. Kegagalan Decision Tree dalam mengenali kelas Rendah menjadi salah satu keterbatasan penelitian ini yang perlu menjadi catatan, mengingat kecilnya jumlah sampel pada kelas tersebut merupakan konsekuensi langsung dari keterbatasan ukuran dataset primer yang dapat dikumpulkan pada populasi UMKM kuliner yang relatif terbatas dan tersebar.

Implementasi Sistem Klasifikasi

Model Random Forest sebagai model dengan performa terbaik selanjutnya diimplementasikan ke dalam aplikasi web interaktif berbasis Streamlit yang di-deploy melalui Hugging Face Spaces. Aplikasi ini memungkinkan pelaku UMKM memasukkan data karakteristik usaha melalui formulir input, kemudian memperoleh hasil klasifikasi tingkat pertumbuhan beserta tingkat keyakinan model pada setiap kelas, penjelasan faktor-faktor utama yang memengaruhi hasil prediksi berbasis feature importance, serta rekomendasi aksi yang disesuaikan dengan kategori hasil. Dengan demikian, penelitian ini tidak hanya menghasilkan model klasifikasi yang akurat, tetapi juga menyediakan sistem yang transparan dan aplikatif sehingga dapat dimanfaatkan secara



langsung oleh pelaku UMKM maupun pemangku kebijakan dalam mendukung pengambilan keputusan pengembangan usaha.

KESIMPULAN

Penelitian ini berhasil membangun model klasifikasi tingkat pertumbuhan UMKM kuliner di Kota Manado menggunakan algoritma Random Forest dan Decision Tree berdasarkan 101 sampel data primer dengan 16 fitur prediktor dan tiga kelas target (Rendah, Sedang, Tinggi). Hasil evaluasi menunjukkan bahwa Random Forest secara konsisten mengungguli Decision Tree pada seluruh metrik berbasis rata-rata tertimbang, dengan akurasi 61,90% berbanding 52,38% dan F1-score weighted 0,5929 berbanding 0,4812, sekaligus menunjukkan stabilitas cross-validation yang lebih baik. Keunggulan ini konsisten dengan karakteristik teoretis Random Forest sebagai model ensemble yang lebih tahan terhadap overfitting dibandingkan pohon keputusan tunggal, khususnya pada dataset berukuran kecil dan distribusi kelas yang tidak seimbang. Melalui pendekatan Explainable AI berbasis SHAP dan LIME, teridentifikasi bahwa `memiliki_target`, `lama_usaha`, `pencatatan_keuangan`, `pendidikan`, dan `aktif_promosi_medsos` merupakan faktor-faktor yang paling konsisten memengaruhi hasil prediksi, baik pada tingkat global maupun pada tingkat sampel individual. Konsistensi pola antara SHAP dan LIME memperkuat validitas interpretasi bahwa kejelasan target usaha, pengalaman berwirausaha, literasi keuangan, dan keaktifan promosi digital menjadi aspek kunci pembeda UMKM kuliner dengan pertumbuhan tinggi dan rendah. Hasil penelitian ini telah diimplementasikan ke dalam aplikasi klasifikasi berbasis web yang transparan dan aplikatif, sehingga dapat langsung dimanfaatkan oleh pelaku UMKM maupun pemangku kebijakan dalam mendukung pengambilan keputusan pengembangan usaha.

SARAN

Penelitian selanjutnya disarankan memperluas ukuran dataset, khususnya pada kelas Rendah yang pada penelitian ini hanya diwakili oleh 21 sampel, karena keterbatasan jumlah data pada kelas tersebut terbukti menjadi penyebab utama Decision Tree gagal mengenalinya sama sekali. Penerapan teknik penanganan ketidakseimbangan kelas, seperti pembobotan kelas atau oversampling yang diterapkan secara ketat di dalam fold pelatihan untuk menghindari kebocoran data, juga perlu dipertimbangkan. Perbandingan dengan algoritma lain seperti XGBoost, LightGBM, atau Support Vector Machine dapat dilakukan untuk memperoleh gambaran yang lebih komprehensif mengenai algoritma yang paling sesuai untuk karakteristik data UMKM kuliner. Sistem klasifikasi yang telah dibangun dapat dikembangkan lebih lanjut dengan fitur pelatihan ulang model secara berkala serta perluasan cakupan wilayah penelitian ke kota-kota lain di Sulawesi Utara. Pendekatan Explainable AI juga dapat diperluas dengan analisis SHAP dependence plot dan interaction plot untuk mengungkap interaksi antarfitur yang lebih kompleks, seperti antara kepemilikan target usaha dan lama usaha, guna menghasilkan rekomendasi kebijakan pembinaan UMKM kuliner yang lebih spesifik.

**DAFTAR PUSTAKA**

- [1] C. Yolanda, U. Hasanah, and F. Zoromi, "2830-7690 peran usaha mikro, kecil dan menengah (umkm) dalam pengembangan ekonomi indonesia 1," vol. 2, no. 3, pp. 170-186, 2024.
- [2] S. Ramadani, D. A. Ramadhani, M. Ikrom, and L. Muda, "Peran Strategis UMKM dalam Mendorong Pertumbuhan Ekonomi Berkelanjutan di Indonesia," 2025.
- [3] A. T. Novitasari, "No Title," vol. 9, no. 2, pp. 184-204, 2022.
- [4] I. Nirwana, R. Nasrah, I. Rahmadani, S. Pengembangan, and P. Inovasi, "STRATEGI INOVASI UMKM: BIDANG KULINER BERBASIS EKONOMI KREATIF UMKM INNOVATION STRATEGY: CULINARY FIELD BASED ON CREATIVE," vol. X, no. 2, pp. 27-34, 2024.
- [5] P. Gissela *et al.*, "DIGITAL TRANSFORMATION OF MANADO ' S CULINARY MSMES : MARKETING ADAPTATION AND CREATIVE LOCAL PRODUCT," vol. 14, no. 4, pp. 4499-4516, 2025.
- [6] I. G. Agung and K. Lestari, "Tantangan akses pembiayaan umkm terhadap lembaga keuangan formal di indonesia," vol. 4, no. 2, pp. 1-11, 2025.
- [7] R. Dafitri, "Sistem Hambatan Berlapis Digitalisasi UMKM Kuliner : Studi Kualitatif Berbasis Kerangka TOE," vol. 21, no. 3, pp. 153-164, 2025.
- [8] H. Gao and G. Kou, "Machine learning in business and finance : a literature review and research opportunities," *Financ. Innov.*, 2024, doi: 10.1186/s40854-024-00629-z.
- [9] A. Nola, S. Kinasih, A. Nur, J. Tri, and N. Salwa, "Comparative analysis of decision tree and random forest classifiers for structured data classification in machine learning," vol. 5, no. 2, pp. 13-24, 2025.
- [10] M. Erkamim, M. Zidni, and E. Widarti, "Komparasi Algoritme Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM," vol. 02, pp. 127-134, 2023.
- [11] M. Haddouchi and A. Berrado, "A survey and taxonomy of methods interpreting random forest models".
- [12] S. Ali, T. Abuhmed, S. El-sappagh, and K. Muhammad, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," *Inf. Fusion*, vol. 99, no. April, p. 101805, 2023, doi: 10.1016/j.inffus.2023.101805.
- [13] A. Nanda and T. Nasution, "Optimasi Deteksi Intrusi Jaringan Menggunakan Hybrid Model Autoencoder dan Random Forest," vol. 7, no. 4, pp. 2349-2360, 2026, doi: 10.47065/bits.v7i4.9309.
- [14] K. T. Refolino and W. I. Rahayu, "ANALISIS KLASSTER RASIO KEUANGAN PT PEGADAIAN DENGAN METODE K-MEANS DAN RANDOM FOREST Berdasarkan kondisi tersebut , penelitian ini diperlukan untuk mengisi celah penelitian dengan mengintegrasikan metode unsupervised learning dan supervised learning dalam," vol. 21, pp. 1-10, 2026.