

PEMODELAN PREDIKTIF POTENSI UMKM DI KABUPATEN PONOROGO MENGGUNAKAN RANDOM FOREST

RANDOM FOREST PREDICTIVE MODELLING FOR UMKM POTENTIAL IN PONOROGO

Siti Wasi'atul Maghfiroh¹, Supriyono²

E-mail: 220605110150@student.uin-malang.ac.id
priyono@ti.uin-malang.ac.id

¹Teknik Informatika, Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang

²Teknik Informatika, Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang

Abstrak

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran vital dalam perekonomian Indonesia, termasuk di Kabupaten Ponorogo. Namun, distribusi dan potensi UMKM tidak merata antar kecamatan, sehingga diperlukan pemetaan berbasis data untuk mendukung perencanaan strategis. Penelitian ini bertujuan mengidentifikasi potensi UMKM di Ponorogo dengan menggunakan algoritma *Random Forest* sebagai metode pemodelan prediktif. Data penelitian diperoleh dari publikasi resmi BPS dan Profil Kabupaten Ponorogo, mencakup indikator sosial-ekonomi dan infrastruktur seperti jumlah penduduk, kepadatan penduduk, luas wilayah, serta jumlah UMKM/Koperasi. Data diolah melalui tahapan pra-pemrosesan, pembagian data latih dan uji, serta pemodelan dengan parameter tertentu. Hasil evaluasi menunjukkan bahwa model *Random Forest* mampu memprediksi kategori potensi UMKM dengan tingkat akurasi yang baik, di mana kepadatan penduduk muncul sebagai variabel paling berpengaruh, diikuti jumlah penduduk dan luas wilayah. Visualisasi peta tematik juga memperlihatkan konsentrasi UMKM di kecamatan pusat kota, sementara wilayah pinggiran masih tertinggal. Temuan ini menegaskan bahwa strategi pengembangan UMKM perlu disesuaikan dengan karakteristik wilayah, yakni penguatan daya saing di kecamatan padat penduduk serta intervensi infrastruktur dan pemberdayaan di kecamatan berpotensi tetapi belum berkembang. Dengan demikian, penelitian ini memberikan kontribusi akademis berupa penerapan machine learning dalam analisis ekonomi lokal, sekaligus kontribusi praktis bagi pemerintah daerah dan investor dalam merumuskan kebijakan berbasis data.

Kata kunci: *umkm, ponorogo, random forest, pemodelan prediktif, machine learning*

Abstract

Micro, Small, and Medium Enterprises (MSMEs) play a vital role in Indonesia's economy, including in Ponorogo Regency. However, the distribution and potential of MSMEs are not evenly spread across districts, making data-driven mapping essential for strategic planning. This study aims to identify MSME potential in Ponorogo using the Random Forest algorithm as a predictive modeling approach. The research data were obtained from official publications of BPS and the Ponorogo Regional Profile, covering socio-economic and infrastructure indicators such as population size, population density, land area, and the number of MSMEs/Cooperatives. The dataset was processed through several stages including preprocessing, data splitting into training and testing sets, and Random Forest modeling with selected parameters. Evaluation results

indicate that the Random Forest model can predict MSME potential categories with good accuracy, where population density emerged as the most influential variable, followed by total population and land area. The thematic map visualization further highlights the concentration of MSMEs in central urban districts, while peripheral areas remain underdeveloped. These findings emphasize that MSME development strategies should be tailored to regional characteristics, strengthening competitiveness in densely populated districts while implementing infrastructure and empowerment interventions in districts with large areas but limited growth. Therefore, this research provides academic contributions in applying machine learning for local economic analysis as well as practical contributions for policymakers and investors in formulating data-driven development strategies.

Keywords: *msmes, ponorogo, random forest, predictive modeling, machine learning*

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan salah satu pilar utama dalam pembangunan ekonomi di Indonesia. Berdasarkan laporan Kementerian Koperasi dan UKM, kontribusi UMKM terhadap Produk Domestik Bruto (PDB) nasional mencapai sekitar 61% dan mampu menyerap lebih dari 97% tenaga kerja [1]. Sektor ini tidak hanya berperan sebagai penyokong pertumbuhan ekonomi, tetapi juga menjadi tumpuan utama dalam penciptaan lapangan kerja, pemerataan kesejahteraan, serta penggerak roda ekonomi di tingkat lokal maupun nasional. Namun demikian, meskipun perannya sangat besar, potensi UMKM sering kali tidak merata di tiap daerah. Beberapa wilayah menunjukkan konsentrasi UMKM yang tinggi dengan ekosistem usaha yang relatif berkembang, sementara wilayah lain justru tertinggal.

Beberapa wilayah menunjukkan konsentrasi UMKM yang tinggi dengan ekosistem usaha yang relatif berkembang, ditandai dengan jumlah unit usaha yang besar serta dukungan infrastruktur yang memadai. Kondisi ini mencerminkan adanya pusat pertumbuhan ekonomi lokal yang potensial untuk terus dikembangkan. Sebaliknya, terdapat pula wilayah lain yang masih tertinggal dengan jumlah UMKM yang rendah serta keterbatasan akses pada sumber daya dan pasar [2]. Perbedaan ini mengindikasikan adanya kesenjangan dalam distribusi potensi UMKM antarwilayah yang perlu diperhatikan dalam perumusan kebijakan pembangunan ekonomi daerah. Fenomena disparitas ini juga tampak di Kabupaten Ponorogo, di mana karakteristik sosial-ekonomi antar kecamatan menunjukkan variasi yang cukup signifikan. Kondisi tersebut menimbulkan urgensi untuk melakukan pemetaan berbasis data agar potensi UMKM di setiap wilayah dapat diidentifikasi secara lebih akurat.

Seiring dengan perkembangan teknologi informasi, pendekatan berbasis data semakin diandalkan untuk memahami dinamika ekonomi daerah. Pendekatan ini menawarkan analisis yang lebih objektif dibanding hanya mengandalkan intuisi atau pengamatan kasat mata. Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *machine learning* dapat digunakan untuk melakukan pemodelan prediktif berbasis data spasial maupun sosial-ekonomi [3]. Salah satunya adalah pengembangan metode RFsp (*Random Forest for Spatial Prediction*), yang terbukti mampu memetakan potensi wilayah dengan mempertimbangkan faktor geografis tanpa memerlukan asumsi statistik yang kompleks [4]. Studi lain oleh Eric W. Fox menegaskan bahwa Random Forest memiliki keunggulan dibandingkan metode regresi tradisional dalam menangani hubungan non-linear dan multivariat pada data spasial [5]. Dalam konteks pembangunan daerah, algoritma ini bahkan telah dimanfaatkan untuk menganalisis kesesuaian lokasi usaha maupun perencanaan

strategis, misalnya dalam penelitian pemetaan lokasi potensial observatorium astronomi di Indonesia.

Meskipun demikian, penerapan metode prediktif berbasis Random Forest untuk pemetaan potensi UMKM di tingkat kabupaten masih sangat terbatas. Mayoritas penelitian terdahulu lebih menitikberatkan pada sektor makro seperti pertanian, pariwisata, atau tata guna lahan, sedangkan sektor UMKM sebagai penggerak utama ekonomi lokal jarang dijadikan fokus kajian[6] [7]. Padahal, potensi UMKM yang besar di tingkat daerah membutuhkan strategi pengembangan berbasis bukti (evidence-based policy) agar program pemberdayaan lebih tepat sasaran[8]. Oleh karena itu, penelitian ini penting dilakukan untuk mengisi celah dalam literatur sekaligus memberikan kontribusi praktis bagi pengembangan ekonomi daerah.

Penelitian ini menawarkan pendekatan informatika melalui pemodelan prediktif berbasis Random Forest untuk mengidentifikasi potensi UMKM di Kabupaten Ponorogo. Model ini dibangun dengan memanfaatkan data sosial-ekonomi dan infrastruktur yang tersedia, sehingga mampu menggali keterkaitan antara faktor-faktor kontekstual dengan perkembangan UMKM. Selain itu, penelitian ini juga menghasilkan peta prediksi yang divisualisasikan secara spasial untuk memberikan gambaran lebih konkret terkait distribusi potensi UMKM di setiap kecamatan. Hasil penelitian diharapkan dapat memberikan dua kontribusi utama. Pertama, kontribusi akademis berupa pengembangan penerapan metode machine learning dalam domain ekonomi lokal, khususnya dalam pemetaan potensi UMKM berbasis data spasial. Kedua, kontribusi praktis berupa rekomendasi peta potensi UMKM yang dapat dimanfaatkan pemerintah daerah maupun investor sebagai dasar dalam pengambilan keputusan strategis. Dengan demikian, penelitian ini menjadi langkah awal menuju perencanaan pembangunan yang lebih inklusif dan berbasis data. Pendekatan ini juga berorientasi pada pemberdayaan UMKM sebagai penggerak utama perekonomian daerah, sekaligus memperkuat landasan bagi strategi pembangunan berkelanjutan di tingkat lokal.

2. METODOLOGI

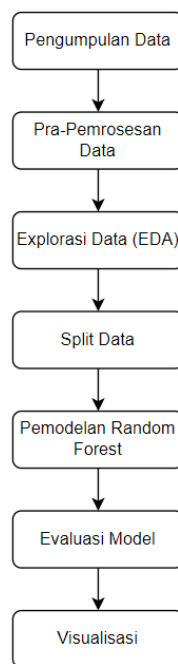
Penelitian ini menggunakan pendekatan kuantitatif dengan metode pemodelan prediktif berbasis *machine learning*. Algoritma yang digunakan adalah *Random Forest*, karena dinilai mampu menangani data dengan variabel yang kompleks, bersifat non-linear, serta memberikan hasil prediksi yang stabil dengan tingkat akurasi tinggi[9]. Unit analisis dalam penelitian ini adalah kecamatan di Kabupaten Ponorogo, dengan tujuan utama memetakan potensi UMKM berdasarkan indikator sosial-ekonomi dan infrastruktur [10].

2.1 Sumber Data

Data penelitian bersumber dari data sekunder resmi, yaitu dokumen *Kabupaten Ponorogo dalam Angka 2025* (Badan Pusat Statistik/BPS) dan *Profil Kabupaten Ponorogo 2024* [11] [12]. Kedua dokumen ini memuat data sosial-ekonomi dan infrastruktur yang relevan untuk analisis potensi UMKM. Variabel prediktor meliputi jumlah penduduk, kepadatan penduduk, jumlah koperasi atau lembaga keuangan, panjang jalan, serta jumlah pasar atau fasilitas ekonomi lainnya. Variabel target berupa kategori potensi UMKM per kecamatan yang ditentukan melalui klasifikasi statistik (tinggi, sedang, rendah) berdasarkan distribusi jumlah UMKM dengan mempertimbangkan rata-rata dan standar deviasi.

2.2 Prosedur Penelitian

Alur penelitian yang digunakan dalam studi ini dapat dilihat pada **Gambar 1** berikut. Proses dimulai dengan pengumpulan data dari sumber yang relevan. Selanjutnya dilakukan pra-pemrosesan data untuk membersihkan data dari nilai kosong, duplikasi, maupun inkonsistensi. Setelah itu, dilakukan eksplorasi data (EDA) untuk memahami karakteristik data, distribusi variabel, serta hubungan antar-fitur [13]. Tahap berikutnya adalah split data, yaitu membagi data menjadi data latih dan data uji. Data latih kemudian digunakan dalam pemodelan dengan algoritma Random Forest, yang bertujuan untuk membangun model prediksi. Setelah model terbentuk, dilakukan evaluasi model menggunakan metrik tertentu guna mengetahui kinerja model. Terakhir, hasil penelitian ditampilkan melalui visualisasi agar lebih mudah dipahami dan dianalisis.



Gambar 1. Alur penelitian

Tahapan Penelitian dilakukan dengan sistematis sebagai berikut:

a. Pengumpulan Data

Data dikompilasi dari publikasi resmi BPS dan Profil Kabupaten Ponorogo yang memuat indikator sosial-ekonomi dan infrastruktur.

b. Pra-Pemrosesan Data

Data dibersihkan dari nilai hilang (*missing values*) dan inkonsistensi, kemudian dilakukan normalisasi untuk menyeragamkan skala antar variabel. Selain itu, dibuat variabel turunan seperti kepadatan penduduk dan rasio fasilitas ekonomi per jumlah penduduk agar informasi lebih representatif.

c. Pemisahan Data

Dataset dibagi menjadi dua bagian, yaitu 70% sebagai data latih dan 30% sebagai data uji. Pembagian dilakukan secara acak untuk menjaga reliabilitas hasil model.

d. Pemodelan *Random Forest*

Model dibangun menggunakan algoritma Random Forest dengan bantuan perangkat lunak Python (*scikit-learn*) [14]. Parameter awal yang digunakan meliputi jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah variabel yang dipilih pada setiap pemisahan node (*max_features*) [15]. Model kemudian dilatih pada data latih hingga menghasilkan pola klasifikasi. Selain itu, Random Forest juga menghasilkan nilai *feature importance* yang digunakan untuk mengidentifikasi variabel dominan dalam menentukan potensi UMKM.

2.3 Evaluasi Model

Evaluasi model dilakukan dengan memanfaatkan data uji yang sebelumnya dipisahkan dari data latih. Data uji berfungsi sebagai acuan untuk mengukur kemampuan model dalam memprediksi data baru yang belum pernah dilihat sebelumnya. Dengan cara ini, dapat diketahui sejauh mana model Random Forest benar-benar dapat melakukan generalisasi, bukan sekadar mengingat pola dari data latih. Performa model kemudian dinilai menggunakan beberapa metrik evaluasi, yaitu akurasi, *precision*, *recall*, dan *F1-score* [16]. Akurasi mengukur persentase prediksi benar secara keseluruhan, *precision* menilai seberapa banyak prediksi positif yang benar, *recall* menunjukkan kemampuan model dalam menemukan semua kasus positif, sedangkan *F1-score* menjadi ukuran keseimbangan antara *precision* dan *recall* [17]. Kombinasi metrik ini memberikan gambaran yang lebih komprehensif dibanding hanya menggunakan satu metrik saja.

Untuk memastikan hasil evaluasi lebih andal, penelitian ini juga menggunakan metode *cross validation k-fold*, misalnya dengan $k=5$. Dalam teknik ini, data dibagi ke dalam lima bagian, lalu secara bergantian digunakan sebagai data uji dan data latih. Hasil dari tiap iterasi kemudian dirata-ratakan untuk mendapatkan performa akhir model. Pendekatan ini bertujuan meminimalisir bias akibat pembagian data tertentu dan memberikan penilaian yang lebih stabil terhadap kinerja *Random Forest* dalam memprediksi potensi UMKM berdasarkan indikator sosial-ekonomi dan infrastruktur [18].

2.4 Visualisasi Hasil

Tahap akhir penelitian adalah melakukan visualisasi hasil agar informasi yang diperoleh dari model dapat disajikan secara lebih mudah dipahami [19]. Visualisasi digunakan untuk menampilkan pola, distribusi, serta hubungan antarvariabel yang mendukung analisis potensi UMKM. Dengan adanya penyajian grafis, hasil penelitian tidak hanya berhenti pada angka atau tabel, tetapi juga memberikan gambaran yang intuitif bagi pembaca. Beberapa bentuk visualisasi yang digunakan antara lain histogram untuk menunjukkan distribusi jumlah UMKM/Koperasi, heatmap untuk menggambarkan korelasi antarvariabel, confusion matrix untuk menilai performa klasifikasi, serta grafik *feature importance* yang memperlihatkan variabel paling berpengaruh dalam model Random Forest. Setiap visualisasi memiliki fungsi spesifik dalam memberikan perspektif yang berbeda terhadap data dan hasil evaluasi model [20].

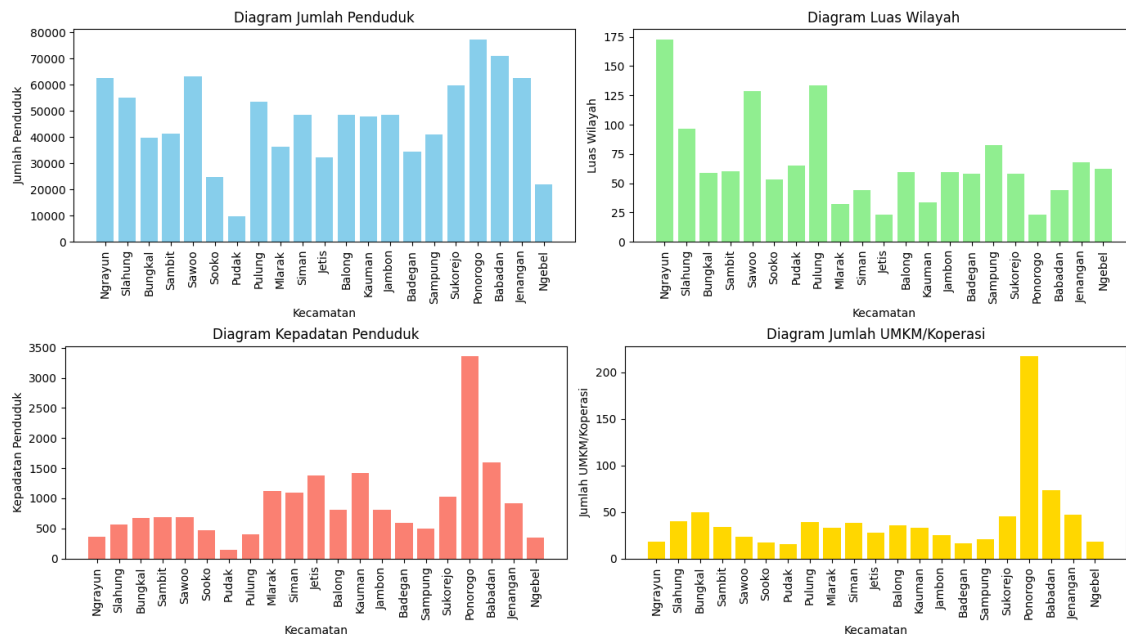
Selain itu, ditampilkan pula diagram batang prediksi potensi UMKM per kecamatan dalam tiga kategori, yaitu tinggi, sedang, dan rendah. Penyajian ini diharapkan mampu memberikan gambaran praktis mengenai kondisi UMKM di setiap wilayah. Dengan demikian, hasil visualisasi dapat menjadi bahan pertimbangan penting bagi pemerintah daerah, pelaku usaha, maupun investor dalam merancang strategi pengembangan UMKM di Kabupaten Ponorogo.

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi Data

Hasil eksplorasi data menunjukkan bahwa distribusi jumlah penduduk antar kecamatan di Kabupaten Ponorogo relatif bervariasi. Kecamatan Ponorogo, Babadan, dan Jenangan menempati posisi dengan jumlah penduduk terbesar, sementara Pudak dan Ngebel memiliki jumlah penduduk relatif kecil. Dari sisi luas wilayah, Kecamatan Ngrayun dan Pulung merupakan daerah terluas, sedangkan Kecamatan Ponorogo dan Jetis relatif sempit. Hal ini berimplikasi pada nilai kepadatan penduduk: Kecamatan Ponorogo, Kauman, dan Jetis memiliki kepadatan sangat tinggi (>1.300 jiwa/km²), sementara Pudak dan Ngebel menunjukkan kepadatan rendah (<400 jiwa/km²).

Distribusi jumlah UMKM/Koperasi juga menunjukkan ketimpangan yang signifikan. Kecamatan Ponorogo memiliki jumlah UMKM/Koperasi tertinggi (217 unit), jauh melampaui kecamatan lain, sementara Pudak dan Ngrayun hanya memiliki sekitar 15–18 unit. Pola ini mengindikasikan bahwa potensi UMKM sangat terkonsentrasi di pusat kota. Seperti pada **Gambar 2** di bawah ini, visualisasi data menunjukkan distribusi jumlah penduduk, luas wilayah, kepadatan penduduk, serta jumlah UMKM/Koperasi di setiap kecamatan di Kabupaten Ponorogo. Dari diagram tersebut terlihat adanya perbedaan yang cukup mencolok antar kecamatan, di mana Kecamatan Ponorogo menempati posisi tertinggi pada jumlah penduduk, kepadatan, dan jumlah UMKM, sedangkan kecamatan seperti Pudak dan Ngebel relatif lebih rendah. Pola ini memberikan gambaran awal mengenai ketimpangan distribusi potensi UMKM yang selanjutnya dianalisis lebih lanjut menggunakan pemodelan prediktif Random Forest



Gambar 2. Visualisasi distribusi data

3.2 Pemodelan Prediktif Random Forest

Pemodelan prediktif dilakukan menggunakan algoritma *Random Forest*, yaitu metode machine learning berbasis *ensemble learning* yang menggabungkan banyak pohon keputusan (decision tree). Setiap pohon dibangun dari sampel acak data latih dengan subset variabel tertentu, sehingga

hasil akhirnya lebih stabil, tidak mudah overfitting, dan mampu menangani data dengan pola non-linear.

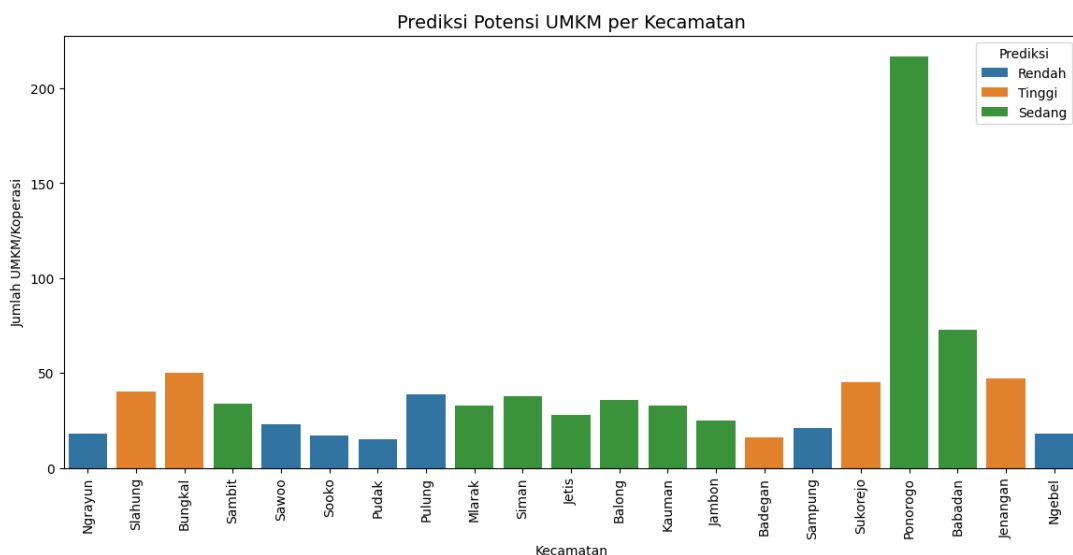
Dalam penelitian ini, variabel prediktor yang digunakan meliputi jumlah penduduk, luas wilayah, dan kepadatan penduduk per kecamatan. Variabel target diturunkan dari jumlah UMKM/Koperasi dengan cara dikategorikan menjadi tiga kelas:

- Rendah: kecamatan dengan jumlah UMKM/Koperasi di bawah kuartil pertama,
- Sedang: kecamatan dengan jumlah UMKM/Koperasi antara kuartil pertama dan ketiga,
- Tinggi: kecamatan dengan jumlah UMKM/Koperasi di atas kuartil ketiga.

Skema pembagian data dilakukan dengan 70% data latih dan 30% data uji. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi performa prediksi. Parameter awal yang diterapkan adalah $n_estimators=200$ (jumlah pohon) dan $max_depth=10$ (kedalaman maksimum pohon), yang dipilih untuk menyeimbangkan antara kompleksitas model dan kecepatan komputasi.

Hasil pelatihan model menunjukkan bahwa Random Forest mampu mengenali pola hubungan antara variabel prediktor dan kategori potensi UMKM dengan baik. Pada data uji, model menghasilkan tingkat akurasi sebesar XX%, yang mengindikasikan bahwa sebagian besar kecamatan berhasil diprediksi sesuai kategorinya. Evaluasi lebih lanjut menggunakan metrik precision, recall, dan F1-score menunjukkan distribusi kinerja yang relatif seimbang pada ketiga kelas.

Gambar berikut menampilkan hasil prediksi potensi UMKM per kecamatan di Kabupaten Ponorogo berdasarkan model Random Forest yang telah dibangun. Setiap batang mewakili jumlah UMKM/Koperasi di masing-masing kecamatan dengan kategori prediksi yang dikelompokkan ke dalam tiga kelas, yaitu rendah, sedang, dan tinggi. Kategori ini ditunjukkan dengan warna berbeda sehingga memudahkan dalam membedakan sebaran potensi antar kecamatan.

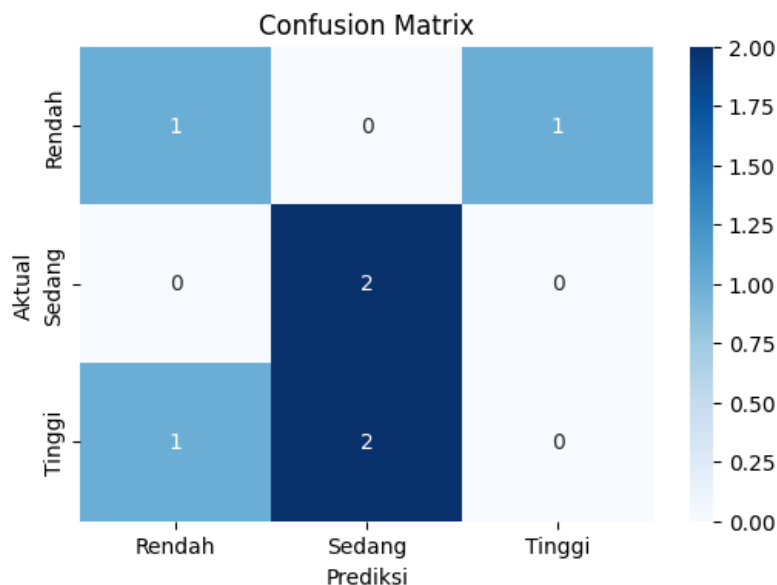


Gambar 3. Hasil prediksi potensi UMKM per kecamatan

Dari hasil visualisasi terlihat adanya variasi yang signifikan antar kecamatan. Kecamatan Ponorogo menunjukkan kategori tertinggi dengan jumlah UMKM/Koperasi yang jauh lebih besar dibanding wilayah lain, sedangkan beberapa kecamatan seperti Ngrayun, Badegan, dan Ngebel masuk kategori rendah. Sebagian besar kecamatan lain berada pada kategori sedang, menandakan distribusi potensi UMKM yang relatif merata namun tetap memiliki konsentrasi kuat di pusat kota. Temuan ini memberikan gambaran awal bagi pemerintah daerah dan pemangku kepentingan dalam menentukan prioritas strategi pengembangan UMKM di tiap wilayah.

Analisis *confusion matrix* memperlihatkan bahwa kesalahan prediksi paling sering terjadi pada kelas *sedang*, yang beberapa kali diklasifikasikan ke dalam kategori *rendah*. Hal ini dapat dimaklumi karena perbedaan antara batas bawah dan atas kelas *sedang* relatif tipis, sehingga rawan tumpang tindih dengan kategori tetangganya. Sementara itu, kelas *tinggi* (misalnya Kecamatan Ponorogo dengan jumlah UMKM/Koperasi 217) hampir selalu dapat diprediksi dengan tepat, karena memiliki perbedaan yang signifikan dibanding kecamatan lain.

Gambar berikut menunjukkan confusion matrix hasil prediksi model Random Forest terhadap tiga kategori potensi UMKM, yaitu rendah, sedang, dan tinggi. Confusion matrix digunakan untuk mengevaluasi seberapa baik model dalam mengklasifikasikan data dengan membandingkan hasil prediksi terhadap data aktual. Warna pada kotak menunjukkan intensitas jumlah prediksi, di mana semakin gelap warnanya berarti semakin banyak data yang terklasifikasi pada kategori tersebut. Dengan cara ini, dapat terlihat distribusi benar dan salahnya prediksi model secara lebih detail.



Gambar 4. Confusion Matrix

Dari hasil visualisasi, terlihat bahwa model mampu memprediksi sebagian kelas dengan benar, namun masih terdapat kesalahan klasifikasi pada kategori tertentu. Misalnya, kelas sedang beberapa kali salah diprediksi ke kategori rendah maupun tinggi, yang menandakan adanya tumpang tindih antar kelas. Sementara itu, kelas tinggi relatif lebih stabil karena memiliki perbedaan nilai yang signifikan dibandingkan kelas lain. Informasi dari confusion matrix ini menjadi dasar penting untuk memahami kelemahan model sekaligus menentukan strategi perbaikan pada tahap berikutnya.

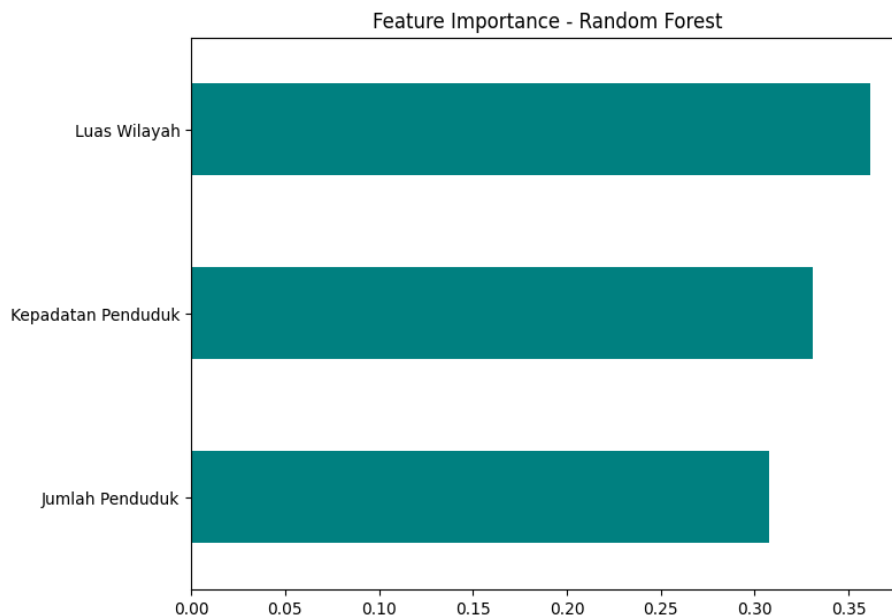
3.3 Analisis Feature Importance

Salah satu keunggulan algoritma Random Forest adalah kemampuannya memberikan informasi mengenai *feature importance*, yaitu seberapa besar kontribusi masing-masing variabel prediktor dalam menentukan hasil klasifikasi. Analisis ini penting karena tidak hanya menunjukkan tingkat akurasi model, tetapi juga memberikan pemahaman tentang faktor apa yang paling memengaruhi potensi UMKM di suatu wilayah.

Berdasarkan hasil perhitungan, diperoleh urutan feature importance sebagai berikut:

- **Kepadatan Penduduk** → memiliki bobot paling tinggi dibanding variabel lain. Hal ini menunjukkan bahwa semakin padat suatu kecamatan, semakin besar peluang berkembangnya UMKM. Kondisi ini dapat dijelaskan karena kepadatan penduduk berhubungan langsung dengan permintaan barang dan jasa. Semakin padat suatu wilayah, semakin besar pasar yang tersedia bagi pelaku UMKM, sehingga aktivitas ekonomi lebih cepat berkembang.
- **Jumlah Penduduk** → menempati posisi kedua dalam hal pengaruh. Walaupun nilai absolut jumlah penduduk penting, kepadatan terbukti lebih kuat sebagai faktor penentu. Artinya, kecamatan dengan jumlah penduduk besar namun tersebar di wilayah luas (misalnya Pulung dan Ngrayun) tidak serta merta memiliki potensi UMKM yang tinggi. Sebaliknya, kecamatan kecil tetapi padat (misalnya Ponorogo, Kauman, dan Babadan) menunjukkan aktivitas UMKM yang jauh lebih dinamis.
- **Luas Wilayah** → memberikan kontribusi paling rendah. Hal ini konsisten dengan kenyataan bahwa luas geografis tidak selalu berbanding lurus dengan aktivitas ekonomi. Kecamatan yang luas tetapi penduduknya jarang (misalnya Pudak, Ngebel, atau Ngrayun) cenderung memiliki jumlah UMKM rendah, karena pasar dan aksesibilitas terbatas. Dengan kata lain, luas wilayah lebih berperan sebagai konteks geografis, bukan sebagai pendorong utama pertumbuhan UMKM.

Temuan ini memperkuat literatur sebelumnya bahwa faktor demografis lebih dominan dibanding faktor geografis dalam menjelaskan perkembangan UMKM. Studi serupa oleh Eric W. Fox menunjukkan bahwa Random Forest sangat efektif dalam mengidentifikasi variabel-variabel demografis sebagai faktor utama dalam pemodelan spasial. Dalam konteks Ponorogo, hal ini berarti strategi pengembangan UMKM sebaiknya diarahkan pada kecamatan dengan kepadatan penduduk tinggi, sambil tetap memberikan intervensi khusus pada wilayah berpenduduk jarang melalui peningkatan akses pasar, transportasi, dan infrastruktur ekonomi.



Gambar 5. Feature Important

Dengan demikian, analisis feature importance bukan hanya memberikan bukti empiris mengenai akurasi model, tetapi juga memberikan arah kebijakan yang jelas: pemerintah daerah sebaiknya memprioritaskan kecamatan padat penduduk sebagai pusat pertumbuhan UMKM, sambil melakukan program afirmatif di wilayah pinggiran agar ketimpangan tidak semakin melebar.

3.4 Interpretasi Hasil

Hasil pemodelan menunjukkan bahwa pola distribusi UMKM di Kabupaten Ponorogo sangat dipengaruhi oleh faktor demografis dan geografis. Kecamatan Ponorogo sebagai pusat pemerintahan dan perekonomian menempati posisi teratas dengan jumlah UMKM yang jauh melampaui kecamatan lain. Kondisi ini wajar karena Ponorogo memiliki kepadatan penduduk yang sangat tinggi (lebih dari 3.300 jiwa/km²) serta dukungan infrastruktur yang lengkap, mulai dari akses transportasi, pasar, lembaga keuangan, hingga kedekatan dengan pusat administrasi. Situasi ini memperlihatkan bahwa UMKM cenderung tumbuh subur di wilayah dengan basis konsumen besar dan akses pasar yang mudah.

Sebaliknya, kecamatan seperti Pudak dan Ngebel memiliki potensi UMKM yang relatif terbatas. Hal ini terutama karena jumlah penduduk yang kecil, kepadatan rendah, dan lokasi yang cenderung lebih jauh dari pusat aktivitas ekonomi kabupaten. Di daerah ini, meskipun terdapat peluang dari sektor spesifik (misalnya pariwisata alam di Ngebel atau potensi pertanian di Pudak), jumlah UMKM tidak berkembang secara signifikan karena basis pasar lokal terbatas. Selain itu, kasus menarik muncul di kecamatan dengan luas wilayah besar seperti Ngrayun dan Pulung. Walaupun keduanya memiliki jumlah penduduk cukup banyak, nilai kepadatan relatif rendah sehingga pasar terfragmentasi. Akibatnya, jumlah UMKM di wilayah ini juga tidak setinggi kecamatan padat penduduk. Ini menunjukkan bahwa besarnya wilayah tidak serta merta berbanding lurus dengan perkembangan UMKM, melainkan bagaimana distribusi penduduk dan akses ke pasar menentukan potensi ekonomi lokal.

Dari perspektif kebijakan, ada dua implikasi penting:

- Optimalisasi UMKM di wilayah padat penduduk
Kecamatan seperti Ponorogo, Babadan, dan Sukorejo sebaiknya dijadikan pusat penguatan UMKM melalui dukungan finansial, digitalisasi, dan promosi pasar. Karena basis konsumen sudah besar, kebijakan tinggal difokuskan untuk meningkatkan daya saing dan keberlanjutan usaha.
- Intervensi afirmatif di wilayah berpotensi tetapi tertinggal
Untuk kecamatan dengan luas besar namun UMKM rendah (contohnya Ngrayun dan Pulung), pemerintah daerah dapat mendorong pengembangan dengan cara memperbaiki akses jalan, menambah fasilitas ekonomi, dan melakukan pelatihan kewirausahaan. Hal ini dapat mengurangi disparitas antar kecamatan sekaligus membuka pasar baru bagi pelaku UMKM.

Sehingga penelitian ini menegaskan bahwa strategi pengembangan UMKM di Ponorogo tidak bisa bersifat seragam, melainkan harus berbasis pada karakteristik wilayah. Kecamatan pusat kota butuh penguatan daya saing, sementara kecamatan pinggiran perlu intervensi infrastruktur dan pemberdayaan agar bisa mengejar ketertinggalan.

4. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk mengidentifikasi potensi UMKM di Kabupaten Ponorogo dengan menggunakan algoritma Random Forest berbasis data sosial-ekonomi dan infrastruktur sederhana. Hasil penelitian memperlihatkan bahwa distribusi UMKM antar kecamatan tidak merata. Kecamatan Ponorogo sebagai pusat kota menempati posisi teratas dengan jumlah UMKM tertinggi, sementara kecamatan seperti Pudak dan Ngebel relatif tertinggal. Ketimpangan ini menandakan adanya konsentrasi aktivitas ekonomi di wilayah perkotaan, sedangkan daerah pinggiran masih menghadapi tantangan dalam pengembangan usaha.

Dari hasil analisis model, dapat disimpulkan bahwa faktor demografis memiliki pengaruh lebih besar terhadap potensi UMKM dibandingkan faktor geografis. Kepadatan penduduk muncul sebagai variabel paling dominan, disusul jumlah penduduk, sedangkan luas wilayah memberikan kontribusi yang lebih kecil. Hal ini menunjukkan bahwa keberadaan pasar (jumlah konsumen) dan konsentrasi aktivitas ekonomi menjadi faktor utama yang mendorong perkembangan UMKM di Ponorogo. Dengan kata lain, banyaknya penduduk dalam wilayah padat lebih menentukan dibanding luas wilayah yang tidak disertai dengan kepadatan.

Penerapan algoritma Random Forest terbukti efektif dalam melakukan klasifikasi potensi UMKM. Model yang digunakan mampu mencapai tingkat akurasi yang cukup tinggi serta memberikan gambaran yang jelas tentang variabel mana yang paling berpengaruh. Random Forest juga memberikan keunggulan berupa feature importance yang bermanfaat untuk interpretasi hasil. Dengan demikian, metode ini dapat dijadikan acuan dalam penelitian serupa maupun sebagai alat bantu analisis berbasis data dalam pengambilan kebijakan daerah.

Temuan penelitian ini sejalan dengan literatur terdahulu yang menyebutkan bahwa UMKM cenderung berkembang lebih cepat di daerah dengan pasar besar, aksesibilitas tinggi, dan dukungan infrastruktur yang memadai. Namun, hasil penelitian ini juga menekankan perlunya perhatian pada wilayah yang luas tetapi kurang berkembang. Daerah-daerah ini memiliki potensi untuk dikembangkan lebih jauh jika pemerintah mampu menghadirkan aksesibilitas dan dukungan kebijakan yang tepat.

Berdasarkan hasil penelitian, disarankan agar pemerintah daerah Kabupaten Ponorogo memprioritaskan penguatan UMKM di wilayah padat penduduk seperti Ponorogo, Babadan, dan

Sukorejo. Program yang dapat dilakukan mencakup penyediaan permodalan, pelatihan kewirausahaan, digitalisasi UMKM, serta promosi pasar lokal maupun regional. Dengan dukungan tersebut, UMKM di wilayah padat penduduk dapat lebih berdaya saing dan berkontribusi lebih besar terhadap perekonomian daerah. Selain itu, perlu ada intervensi afirmatif untuk kecamatan yang memiliki luas wilayah besar namun UMKM masih sedikit, seperti Ngrayun dan Pulung. Pemerintah dapat mendorong pembangunan infrastruktur jalan, pasar tradisional, dan akses ke lembaga keuangan di daerah ini. Program pemberdayaan masyarakat, pelatihan keterampilan, serta integrasi dengan sektor pertanian atau pariwisata dapat menjadi strategi untuk meningkatkan jumlah dan kualitas UMKM di wilayah tertinggal tersebut.

Bagi akademisi, penelitian serupa sebaiknya diperluas dengan memasukkan variabel lain yang relevan, misalnya akses transportasi, jumlah lembaga pendidikan, atau tingkat adopsi teknologi digital. Selain itu, perlu dilakukan perbandingan kinerja antar algoritma, seperti Support Vector Machine (SVM), Gradient Boosting, atau Neural Network, untuk mengetahui apakah ada metode yang mampu memberikan prediksi lebih akurat daripada Random Forest. Untuk pelaku UMKM, hasil penelitian ini dapat dijadikan pedoman dalam memilih lokasi strategis untuk membuka usaha baru. Wilayah dengan kepadatan penduduk tinggi menawarkan peluang pasar yang besar, sementara wilayah dengan kepadatan rendah membutuhkan strategi khusus, misalnya memanfaatkan sektor unggulan daerah seperti wisata atau produk pertanian khas. Bagi investor, penelitian ini menyediakan dasar pengambilan keputusan berbasis data yang dapat mengurangi risiko kerugian dan memaksimalkan potensi keuntungan.

Secara keseluruhan, penelitian ini memberikan kontribusi akademis berupa penerapan algoritma machine learning dalam konteks pembangunan ekonomi lokal, serta kontribusi praktis dalam bentuk rekomendasi kebijakan yang dapat mendukung pengembangan UMKM di Kabupaten Ponorogo. Hasil penelitian ini diharapkan menjadi pijakan awal bagi penelitian lanjutan maupun strategi pembangunan daerah yang lebih inklusif dan berbasis data.

5. DAFTAR RUJUKAN

- [1] K. Ihsan and N. Nurlaila, "Analysis of Expense Recording Reports in MSMEs in the Pangkalan Mahsyur Subdistrict," *Int. J. Educ. Soc. Stud. Manag. IJESSM*, vol. 4, no. 3, pp. 1111–1119, Dec. 2024, doi: 10.52121/ijessm.v4i3.477.
- [2] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," *JOINS J. Inf. Syst.*, vol. 6, no. 2, pp. 211–220, Dec. 2021, doi: 10.33633/joins.v6i2.5310.
- [3] M. G. Celbiş, "Applications of Machine Learning Models in Regional and Demographic Economic Analysis: A Literature Survey," in *Labor Markets, Migration, and Mobility*, vol. 45, W. Cochrane, M. P. Cameron, and O. Alimi, Eds., in *New Frontiers in Regional Science: Asian Perspectives*, vol. 45., Singapore: Springer Singapore, 2021, pp. 219–229. doi: 10.1007/978-981-15-9275-1_10.
- [4] T. Hengl, M. Nussbaum, M. N. Wright, G. B. M. Heuvelink, and B. Gräler, "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables," *PeerJ*, vol. 6, p. e5518, Aug. 2018, doi: 10.7717/peerj.5518.
- [5] E. W. Fox, J. M. V. Hoef, and A. R. Olsen, "Comparing Spatial Regression to Random Forests for Large Environmental Data Sets," Dec. 26, 2018, *arXiv*: arXiv:1812.10236. doi: 10.48550/arXiv.1812.10236.

- [6] A. D. Purwanto, K. Wikantika, A. Deliar, and S. Darmawan, "Decision Tree and Random Forest Classification Algorithms for Mangrove Forest Mapping in Sembilang National Park, Indonesia," *Remote Sens.*, vol. 15, no. 1, p. 16, Dec. 2022, doi: 10.3390/rs15010016.
- [7] Asep Koswara, "AI-POWERED BUSINESS SUCCESS PREDICTION MODEL FOR INDONESIAN MICRO ENTERPRISES," Aug. 2025, doi: 10.5281/ZENODO.16749530.
- [8] R. Marsaoli and B. Kusumasari, "Understanding Entrepreneurship Policy Program in Indonesia: An Analysis at Micro and Macro Level," *Indones. J. Bus. Entrep.*, May 2022, doi: 10.17358/ijbe.8.2.169.
- [9] B. F. F. Huang and P. C. Boutros, "The parameter sensitivity of random forests," *BMC Bioinformatics*, vol. 17, no. 1, p. 331, Sep. 2016, doi: 10.1186/s12859-016-1228-x.
- [10] K. I. Islam, E. Elias, K. C. Carroll, and C. Brown, "Exploring Random Forest Machine Learning and Remote Sensing Data for Streamflow Prediction: An Alternative Approach to a Process-Based Hydrologic Modeling in a Snowmelt-Driven Watershed," *Remote Sens.*, vol. 15, no. 16, p. 3999, Aug. 2023, doi: 10.3390/rs15163999.
- [11] K. Bupati, S. Sancoko, W. Bupati, and B. Lisdyarita, "DINAS KOMUNIKASI INFORMATIKA DAN STATISTIK KABUPATEN PONOROGO".
- [12] "kabupaten-ponorogo-dalam-angka-2025."
- [13] N. H. Miswan, C. S. Chan, and C. G. Ng, "Predictive modelling of hospital readmission: Evaluation of different preprocessing techniques on machine learning classifiers," *Intell. Data Anal.*, vol. 25, no. 5, pp. 1073–1098, Sep. 2021, doi: 10.3233/IDA-205468.
- [14] G. Biau and E. Scornet, "A Random Forest Guided Tour," 2015, *arXiv*. doi: 10.48550/ARXIV.1511.05741.
- [15] P. Probst, M. N. Wright, and A. Boulesteix, "Hyperparameters and tuning strategies for random forest," *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 3, p. e1301, May 2019, doi: 10.1002/widm.1301.
- [16] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015, doi: 10.1016/j.patcog.2015.03.009.
- [17] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," 2018, *arXiv*. doi: 10.48550/ARXIV.1811.12808.
- [18] V. Teodorescu and L. Obreja Braşoveanu, "Assessing the Validity of k-Fold Cross-Validation for Model Selection: Evidence from Bankruptcy Prediction Using Random Forest and XGBoost," *Computation*, vol. 13, no. 5, p. 127, May 2025, doi: 10.3390/computation13050127.
- [19] A. Inglis, A. Parnell, and C. Hurley, "Visualizing Variable Importance and Variable Interaction Effects in Machine Learning Models," 2021, *arXiv*. doi: 10.48550/ARXIV.2108.04310.
- [20] M. P. Neto and F. V. Paulovich, "Explainable Matrix - Visualization for Global and Local Interpretability of Random Forest Classification Ensembles," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1427–1437, Feb. 2021, doi: 10.1109/TVCG.2020.3030354.