Journal of Informatics and Advanced Computing (JIAC)
Vol. 6 No.2, November 2025
E-ISSN : 2722-0346

Komparasi Algoritma Machine Learning untuk Prediksi Lokasi Strategis UMKM Berbasis Data Spasial dan Sosial-Ekonomi

Rizki Hesananda[-1] , Kevin Anugrah Putra Hasibuan[-2]
[1]Teknologi Informasi / Universitas Siber Indonesia, [2]Teknik Informatika / Universitas Krisnadwipayana
hessananda@cyber-univ.ac.id[1], kevin.hasibuan@gmail.com[2]

Abstrak — Penentuan lokasi usaha yang strategis merupakan faktor krusial bagi keberhasilan Usaha Mikro, Kecil, dan Menengah (UMKM). Penelitian ini membahas analisis komparatif tiga algoritma machine learning, yaitu Random Forest, K-Nearest Neighbors (KNN), dan XGBoost, untuk memprediksi kelayakan lokasi UMKM berdasarkan data spasial dan sosial-ekonomi. Dataset terdiri dari 500 sampel dengan lima fitur utama, termasuk jumlah pesaing, jarak ke jalan raya, kepadatan penduduk, dan biaya sewa, serta ditemukan ketidakseimbangan kelas pada variabel target 'lokasi strategis' . Metodologi penelitian mencakup imputasi nilai hilang, standardisasi fitur, rekayasa fitur rasio penduduk–pesaing, serta optimasi hyperparameter menggunakan GridSearchCV. Hasil eksperimen menunjukkan bahwa ketiga model mencapai akurasi tinggi (≈90%), tetapi memiliki keterbatasan dalam memprediksi kelas minoritas, ditunjukkan oleh nilai precision dan F1-score yang rendah untuk kelas "Strategis" . Analisis feature importance menegaskan bahwa jumlah pesaing merupakan faktor paling berpengaruh, disusul kepadatan penduduk dan rasio penduduk–pesaing. Penelitian ini menyimpulkan bahwa meskipun model mampu melakukan klasifikasi dasar dengan baik, penanganan ketidakseimbangan kelas dan penambahan fitur relevan sangat diperlukan untuk meningkatkan akurasi prediksi lokasi strategis UMKM. Temuan ini memberikan kontribusi dalam pengembangan sistem rekomendasi lokasi berbasis data untuk pengambilan keputusan UMKM.

Kata Kunci: UMKM, KNN, XG-Boost, Data Spasial

I. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan sektor strategis yang memainkan peran penting dalam pertumbuhan ekonomi nasional. UMKM berkontribusi terhadap penciptaan lapangan kerja, peningkatan pendapatan masyarakat, dan pemerataan kegiatan ekonomi di berbagai wilayah. Namun demikian, tantangan mendasar yang dihadapi pelaku UMKM adalah menentukan lokasi usaha yang benar-benar strategis [1], [2]. Pemilihan lokasi berdampak langsung pada visibilitas usaha, potensi pasar, aksesibilitas konsumen, serta tingkat persaingan [3]. Oleh karena itu, proses penentuan lokasi tidak dapat hanya mengandalkan intuisi, melainkan membutuhkan pendekatan analitis berbasis data [4]. Kebutuhan akan metode yang lebih objektif dalam mengevaluasi lokasi menjadi semakin mendesak seiring meningkatnya kompleksitas lingkungan bisnis.

Dalam konteks ini, pemanfaatan machine learning menghadirkan peluang baru untuk mengolah data spasial dan non-spasial secara komprehensif. Machine learning memungkinkan identifikasi pola dan hubungan antarvariabel yang sulit diamati melalui metode konvensional [5], [6]. Dengan kemampuan memodelkan berbagai skenario dan memberikan prediksi berbasis statistik, teknologi ini dapat mendukung pengambilan keputusan strategis dalam memilih lokasi UMKM. Pemanfaatan model prediktif juga membuka ruang bagi evaluasi yang lebih akurat terkait tingkat persaingan, daya beli masyarakat, biaya operasional, dan faktor-faktor eksternal lain yang berpotensi mempengaruhi keberhasilan usaha [7], [8]. Dalam penelitian ini, pendekatan machine learning digunakan sebagai fondasi analitis dalam menilai kelayakan lokasi usaha.

Dataset yang digunakan dalam penelitian berisi 500 sampel lokasi UMKM yang mencakup berbagai aspek penting, seperti jumlah pesaing, jarak ke jalan raya, kepadatan penduduk, biaya sewa, dan kategorisasi lokasi sebagai strategis atau tidak strategis. Namun, dataset tersebut juga memiliki tantangan berupa nilai yang hilang pada seluruh fitur serta ketidakseimbangan kelas yang signifikan, di mana lokasi tidak strategis jauh lebih dominan dibandingkan lokasi strategis. Karakteristik dataset semacam ini menuntut proses persiapan data yang tepat untuk memastikan model dapat mempelajari pola secara optimal dan tidak bias terhadap kelas mayoritas [9].

Sejalan dengan tujuan penelitian, tiga algoritma machine learning digunakan untuk melakukan perbandingan performa, yaitu Random Forest, K-Nearest Neighbors (KNN), dan XGBoost. Ketiga algoritma ini dipilih karena kemampuannya dalam menangani data kompleks dan menghasilkan performa prediksi yang kuat [10], [11]. Random Forest dan XGBoost memberikan keunggulan melalui mekanisme ensemble yang mampu meningkatkan stabilitas prediksi, sementara KNN menawarkan pendekatan non-parametrik yang berguna dalam klasifikasi berbasis kemiripan data [12], [13], [14]. Pemilihan dan evaluasi tiga algoritma ini memberikan kontribusi penting dalam memahami model mana yang paling efektif untuk kasus penentuan lokasi UMKM.

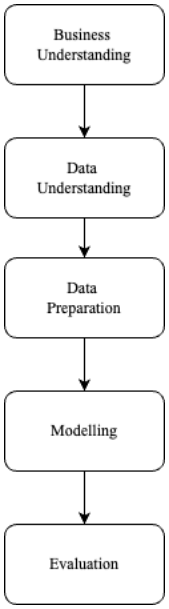
Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, F1-score, dan analisis kurva ROC. Meskipun ketiga model menunjukkan akurasi tinggi pada kisaran 90–91%, evaluasi yang lebih mendalam mengungkapkan kelemahan dalam memprediksi kelas minoritas, khususnya

kategori “Strategis” [15], [16], [17]. Nilai precision dan F1-score yang rendah menunjukkan bahwa model kesulitan mengenali karakteristik lokasi strategis dengan presisi yang baik [18], [19], [20]. Kondisi ini menunjukkan perlunya penanganan ketidakseimbangan kelas serta eksplorasi fitur yang lebih kaya untuk meningkatkan performa model.

Berdasarkan latar belakang tersebut, penelitian ini tidak hanya berfokus pada pengukuran performa algoritma, tetapi juga pada pemahaman mendalam mengenai faktor-faktor yang mempengaruhi kelayakan lokasi. Analisis feature importance dari Random Forest dan XGBoost mengidentifikasi bahwa jumlah pesaing merupakan faktor dominan, disusul oleh kepadatan penduduk serta rasio penduduk-pesaing. Temuan ini menjadi dasar bagi pengembangan sistem rekomendasi lokasi berbasis data dan berpotensi diadopsi oleh pelaku UMKM, pemerintah daerah, maupun pengembang sistem informasi. Dengan demikian, penelitian ini memberikan kontribusi teoretis dan praktis dalam pengambilan keputusan berbasis machine learning untuk penentuan lokasi strategis UMKM.

II. METODE

Penelitian ini menggunakan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai metodologi utama. CRISP-DM dipilih karena menyediakan alur komprehensif untuk memahami permasalahan bisnis, mempersiapkan data, membangun model, hingga mengevaluasi performanya. Tahapan berikut disusun berdasarkan implementasi yang dilakukan dalam laporan.



Gambar 1. Metode Penelitian

Berikut penjelasan mengenai langkah penelitian yang ditampilkan pada Gambar 1:

1. Business Understanding

Tahap ini berfokus pada pemahaman masalah yang ingin diselesaikan, yaitu penentuan lokasi strategis untuk UMKM. Lokasi usaha diklasifikasikan menjadi dua kategori, yaitu Strategis dan Tidak Strategis. Tujuan utama penelitian adalah mengembangkan model prediksi yang mampu mengidentifikasi kelayakan lokasi berdasarkan faktor seperti jumlah pesaing, jarak ke jalan raya, kepadatan penduduk, dan biaya sewa. Keluaran analitis ini diharapkan dapat membantu UMKM dalam mengambil keputusan lokasi yang lebih terukur dan berbasis data.

2. Data Understanding

Dataset terdiri dari 500 sampel dengan empat fitur numerik dan satu fitur kategorikal (lokasi strategis). Data dieksplorasi untuk melihat distribusi fitur, nilai unik, jumlah nilai hilang, dan korelasi antarvariabel. Visualisasi berupa histogram, heatmap korelasi, dan countplot digunakan untuk memahami pola dasar dalam data. Hasil eksplorasi menunjukkan adanya nilai hilang pada seluruh fitur serta ketidakseimbangan kelas, di mana lokasi “Tidak Strategis” jauh lebih banyak dibandingkan lokasi “Strategis”. Temuan ini menjadi landasan untuk tahap persiapan data.

3. Data Preparation

Tahap ini mencakup proses pembersihan dan transformasi data agar siap digunakan untuk pemodelan. Beberapa langkah yang dilakukan meliputi Imputasi nilai hilang menggunakan mean untuk fitur numerik dan modus untuk fitur kategorikal. Encoding fitur kategorikal lokasi_strategis menjadi nilai biner (0 = Tidak, 1 = Ya). Standarisasi fitur numerik menggunakan StandardScaler agar skala antarvariabel seragam. Rekayasa fitur (feature engineering) dengan membuat variabel baru rasio_penduduk_pesaing untuk menangkap hubungan antara potensi pasar dan tingkat persaingan. Pemecahan data menjadi data latih (80%) dan data uji (20%). Langkah-langkah ini memastikan model dapat belajar dari data yang bersih, terstruktur, dan memiliki representasi fitur yang lebih kaya.

4. Modeling

Tiga algoritma klasifikasi diterapkan: Random Forest, K-Nearest Neighbors (KNN), dan XGBoost. Pemilihan algoritma mempertimbangkan kemampuan masing-masing dalam menangani data tabular dan fitur yang bervariasi. Untuk meningkatkan performa, dilakukan tuning hyperparameter menggunakan GridSearchCV dengan validasi silang 5-fold. Parameter penting yang dituning antara lain jumlah pohon (n_estimators) dan kedalaman pohon (max_depth) pada Random Forest, jumlah tetangga pada KNN, serta learning rate pada XGBoost. Selain itu,

model juga divisualisasikan menggunakan kurva ROC untuk menilai kemampuan diskriminatif setiap algoritma.

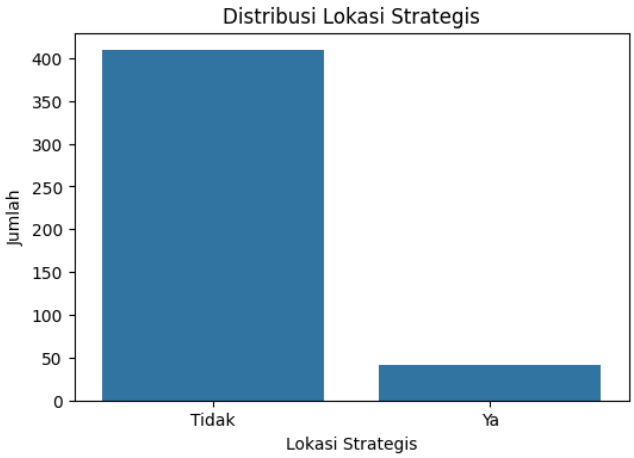
5. Evaluation

Tahap evaluasi menggunakan lima metrik utama: akurasi, precision, recall, F1-score, dan AUC-ROC. Hasil pengujian menunjukkan bahwa ketiga model mencapai akurasi tinggi (≈90%), tetapi kurang optimal dalam memprediksi kelas "Strategis". Hal ini tercermin dari nilai precision dan F1-score yang rendah akibat ketidakseimbangan kelas dalam dataset. Analisis confusion matrix digunakan untuk mengidentifikasi pola kesalahan prediksi. Selain itu, feature importance Random Forest dan XGBoost dianalisis untuk memahami kontribusi fitur terhadap keputusan model.

III. HASIL DAN PEMBAHASAN

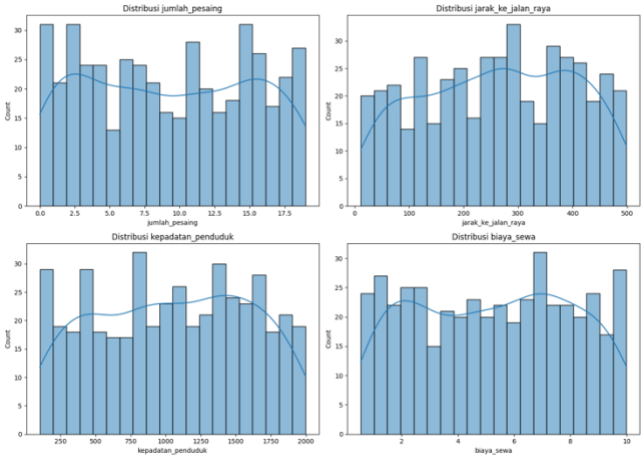
Tahap *Business Understanding* menghasilkan pemahaman bahwa permasalahan utama dalam penelitian ini adalah kebutuhan untuk mengidentifikasi lokasi strategis bagi UMKM berdasarkan variabel spasial dan sosial-ekonomi. Pelaku UMKM kerap mengambil keputusan lokasi secara intuitif, sehingga risiko ketidaktepatan lokasi cukup tinggi. Dengan membangun model klasifikasi yang mengkategorikan lokasi sebagai "Strategis" atau "Tidak Strategis", penelitian ini bertujuan menyediakan dasar pengambilan keputusan yang lebih terukur. Analisis awal menunjukkan bahwa faktor seperti tingkat persaingan, kepadatan penduduk, dan biaya sewa merupakan elemen yang harus dipertimbangkan secara simultan untuk menentukan potensi suatu lokasi. Oleh karena itu, pemanfaatan algoritma machine learning dianggap tepat untuk menemukan pola kelayakan lokasi berdasarkan data historis.

Tahap *Data Understanding* menunjukkan bahwa dataset berisi 500 sampel lokasi dengan empat fitur numerik dan satu variabel target kategorikal. Eksplorasi distribusi data menunjukkan ketidakseimbangan kelas yang signifikan, di mana mayoritas sampel dikategorikan sebagai lokasi "Tidak Strategis" seperti pada Gambar 2.



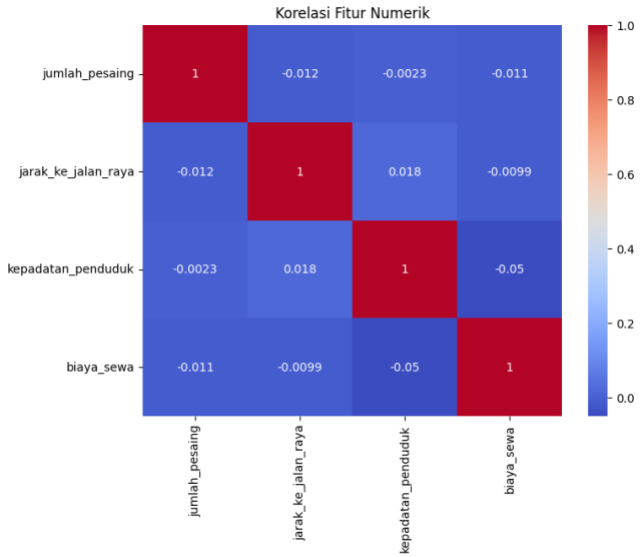
Gambar 2. Distribusi Lokasi

Kondisi ini berpotensi membuat model bias terhadap kelas mayoritas. Selain itu, ditemukan nilai hilang pada seluruh fitur, sehingga proses imputasi menjadi langkah penting dalam tahap persiapan data.



Gambar 3. Histogram Fitur Dataset

Visualisasi histogram pada Gambar 3 mengungkap variasi sebaran pada fitur numerik, dan heatmap korelasi pada Gambar 4 menunjukkan bahwa hubungan linier antar fitur relatif lemah. Hal ini mengindikasikan perlunya model yang mampu menangkap hubungan non-linear dalam data, seperti Random Forest dan XGBoost.

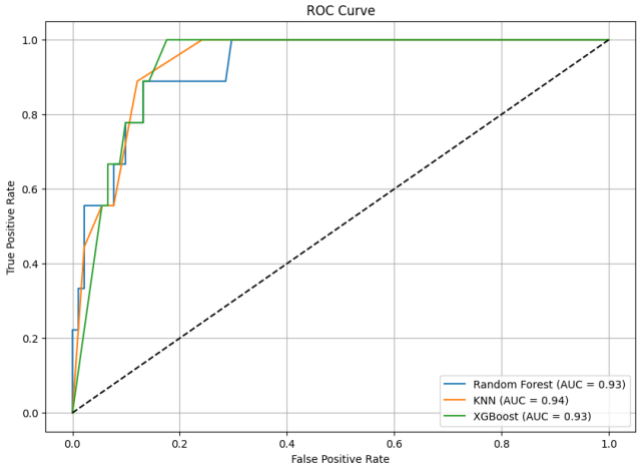


Gambar 4. Heatmap Korelasi Antar Variabel

Pada tahap *Data Preparation*, seluruh nilai hilang berhasil ditangani menggunakan imputasi mean untuk fitur numerik dan modus untuk fitur kategorikal. Variabel target dikodekan menjadi biner agar kompatibel dengan algoritma klasifikasi. Fitur numerik kemudian distandarisasi untuk menyamakan skala antar fitur, terutama penting untuk algoritma berbasis jarak seperti KNN. Penelitian juga melakukan rekayasa fitur dengan menambahkan rasio penduduk terhadap pesaing sebagai indikator tambahan yang

merepresentasikan keseimbangan antara potensi pasar dan intensitas kompetisi. Setelah seluruh proses transformasi, data dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk memastikan evaluasi yang objektif.

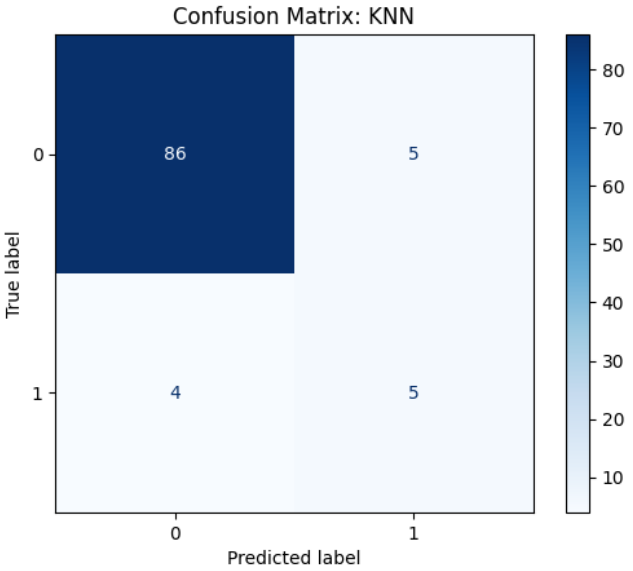
Tahap *Modeling* melibatkan tiga algoritma, yaitu Random Forest, KNN, dan XGBoost. Masing-masing algoritma dioptimalkan menggunakan GridSearchCV dengan validasi silang untuk memastikan pemilihan kombinasi hyperparameter terbaik. Random Forest dibangun dengan variasi jumlah pohon dan kedalaman pohon, sedangkan KNN diuji menggunakan beberapa variasi jumlah tetangga dan metode pembobotan.



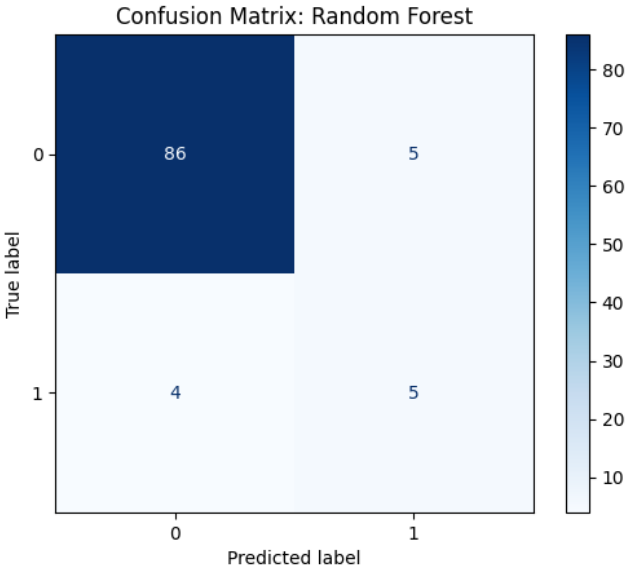
Gambar 5. Kurva ROC setiap Algoritma

XGBoost, sebagai algoritma boosting yang lebih kompleks, diuji dengan beberapa kombinasi *learning rate*, jumlah estimators, dan kedalaman pohon. Ketiga model kemudian dievaluasi menggunakan kurva ROC seperti pada Gambar 5, dan hasilnya menunjukkan bahwa seluruh model memiliki kemampuan diskriminatif yang baik dengan nilai AUC di atas 0,90, meskipun KNN sedikit lebih unggul dibanding dua model lainnya.

Tahap *Evaluation* menampilkan hasil evaluasi menggunakan akurasi, precision, recall, dan F1-score. Meskipun ketiga model menghasilkan akurasi tinggi (sekitar 90–91%), analisis lebih mendalam menunjukkan bahwa performa pada kelas minoritas "Strategis" masih kurang memuaskan. Precision model berada pada kisaran 0,45–0,50, yang berarti separuh prediksi "Strategis" adalah prediksi salah. Recall berada pada angka 0,56, menunjukkan bahwa model hanya berhasil menangkap sekitar setengah dari seluruh lokasi yang sebenarnya strategis. Nilai F1-score yang rendah mengindikasikan bahwa model masih kesulitan melakukan generalisasi terhadap kelas minoritas.



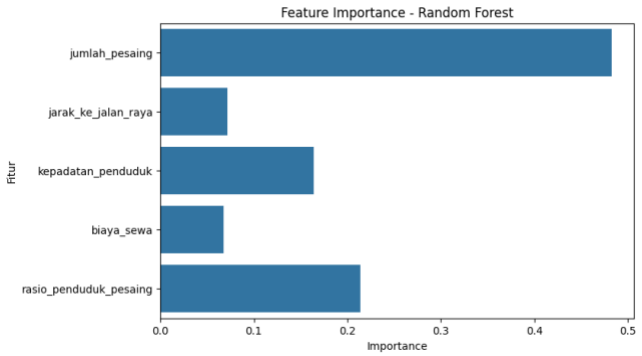
Gambar 6. Kurva ROC setiap Algoritma



Gambar 7. Kurva ROC setiap Algoritma

Confusion matrix pada seluruh model di Gambar 6 dan Gambar 7 memperlihatkan kecenderungan kuat untuk memprediksi kelas mayoritas, yang merupakan konsekuensi langsung dari ketidakseimbangan data.

 E-ISSN : 2722-0346



Gambar 8. *Feature Importance*

Analisis *feature importance* pada Gambar 8 menunjukan Random Forest dan XGBoost menunjukkan bahwa jumlah pesaing merupakan fitur paling dominan dalam mempengaruhi prediksi model. Fitur kepadatan penduduk dan rasio penduduk terhadap pesaing juga memberikan kontribusi signifikan, sedangkan biaya sewa dan jarak ke jalan raya cenderung memiliki pengaruh yang lebih kecil. Penelitian juga menggunakan analisis SHAP untuk memvisualisasikan kontribusi fitur pada level individu, dan hasilnya menunjukkan bahwa nilai rasio penduduk-pesaing yang rendah cenderung menurunkan peluang suatu lokasi dianggap strategis. Sementara itu, nilai kepadatan penduduk yang lebih tinggi cenderung meningkatkan peluang tersebut. Temuan ini memperkuat pemahaman bahwa kelayakan lokasi UMKM sangat dipengaruhi oleh interaksi antara jumlah pesaing dan potensi pasar.

IV. KESIMPULAN

Penelitian ini menyajikan analisis komparatif tiga algoritma machine learning yaitu Random Forest, K-Nearest Neighbors (KNN), dan XGBoost—untuk memprediksi kelayakan lokasi strategis UMKM berdasarkan data spasial dan sosial-ekonomi. Berdasarkan hasil pengolahan data, ketiga model menunjukkan akurasi yang tinggi pada kisaran 90–91%. Namun, akurasi tersebut tidak sepenuhnya mencerminkan performa yang sebenarnya, mengingat dataset memiliki ketidakseimbangan kelas yang signifikan. Model terbukti lebih mampu mengenali kelas mayoritas "Tidak Strategis" dibandingkan kelas minoritas "Strategis", yang menyebabkan precision dan F1-score untuk kelas positif berada pada kategori rendah. Temuan ini mengindikasikan bahwa performa model dalam konteks data tidak seimbang membutuhkan pendekatan tambahan agar prediksi kelas minoritas menjadi lebih akurat.

Analisis feature importance menunjukkan bahwa variabel *jumlah_pesaing* menjadi faktor paling dominan dalam mempengaruhi prediksi model. Hal ini menegaskan bahwa intensitas persaingan merupakan indikator penting dalam menentukan kelayakan lokasi UMKM. Fitur *kepadatan_penduduk* dan fitur rekayasa *rasio_penduduk_pesaing* juga memiliki kontribusi signifikan terhadap keputusan model, menunjukkan adanya hubungan relevan antara potensi pasar, intensitas kompetisi, dan kelayakan lokasi. Sebaliknya, variabel seperti *biaya_sewa* dan *jarak_ke_jalan_raya* memiliki pengaruh yang lebih kecil dalam menentukan prediksi, meskipun tetap berperan sebagai bagian dari konteks lokasi secara keseluruhan.

Dari hasil simulasi deployment, model cenderung memprediksi seluruh sampel baru sebagai lokasi "Tidak Strategis". Meskipun ini sesuai dengan distribusi data yang tidak seimbang, prediksi tersebut memberikan sinyal bahwa model kurang sensitif dalam mendeteksi potensi lokasi strategis. Untuk itu, hasil ini harus ditafsirkan secara hati-hati dan sebaiknya tidak dijadikan dasar tunggal dalam pembuatan keputusan bisnis. Model tetap dapat berfungsi sebagai alat bantu analitis yang memberikan gambaran awal mengenai faktor kelayakan lokasi, namun validasi lapangan dan penilaian kualitatif tetap perlu dilakukan untuk memperoleh rekomendasi lokasi yang lebih komprehensif.

Secara umum, penelitian ini menunjukkan bahwa metode CRISP-DM memberikan alur kerja yang terstruktur dan efektif dalam membangun model prediksi lokasi UMKM. Namun, kualitas prediksi sangat bergantung pada kualitas serta keseimbangan data, pemilihan fitur, dan penerapan teknik evaluasi yang sensitif terhadap ketidakseimbangan. Penelitian ini memberikan landasan awal yang kuat bagi pengembangan sistem rekomendasi lokasi berbasis machine learning, tetapi masih terbuka ruang untuk penyempurnaan model agar lebih akurat dan representatif terhadap kondisi lapangan.

DAFTAR PUSTAKA

[1] N. M. M. Y. Ni Made Marta Yani, Agus Khazin Fauzi, and Ni Nyoman Yuliati, "Pengaruh Pemanfaatan Teknologi Informasi, Penggunaan Informasi Akuntansi Dan Pelatihan Terhadap Kinerja Umkm Di Kecamatan Mataram," *Jurnal Akuntansi dan Keuangan Syariah - ALIANSI*, vol. 6, no. 1, pp. 13–24, 2021, doi: 10.54712/aliansi.v6i1.188.

[2] B. Harto, M. Marliana, P. Pramuditha, and T. Sumarni, "Pemanfaatan Teknologi Informasi Digital , Digital Marketing , dalam Mendorong Sustainability Competitive Bisnis UMKM Kue SuguhWangi di Desa Melatiwangi Kecamatan Cilengkrang Kabupaten Bandung," *Ithabdimas*, vol. 8, no. 1, pp. 221–229, 2024.

[3] A. P. Putra, E. A. Bachtiar, R. Hidayatulloh, A. S. Ramadhani, K. Ummah, and W. Sholihah, "Perancangan Sistem Rekomendasi Komoditas Pertanian Berdasarkan Lokasi Geografis Untuk Meningkatkan Produktivitas Petani," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3936.

[4] N. Syifa and R. N. Fahmi, "Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat," *JOINS (Journal of Information System)*, vol. 6, no. 2, pp. 211–220, 2021, doi: 10.33633/joins.v6i2.5310.

[5] R. Hidayat *et al.*, "IMPLEMENTASI ALGORITMA MACHINE LEARNING MENGGUNAKAN MODEL RANDOM FOREST UNTUK KLASIFIKASI

OBESITAS," *ejournal.itn.ac.id*, vol. 10, no. 1, 2025, doi: 10.51717/simkom.v10i1.703.

[6] D. Kiswanto, F. Ramadhani, N. M. Surbakti, N. A. Nasution, F. Matematika, and I. P. Alam, "Pengembangan dan implementasi sistem deteksi serangan DDoS berbasis algoritma Random Forest," *journal.fkpt.org*, vol. 6, no. 3, pp. 247–256, 2025, doi: 10.47065/bit.v5i2.2203.

[7] S. Lumbantoruan, A. Duha, T. P. Astuti, A. J. Wahidin, B. Rahmatullah, and I. Kurniawati, "Prediksi Harga Saham PT Telkom Indonesia Tbk (TLKM) Menggunakan Linear Regression, Random Forest, dan XGBoost," *journal.ilmudata.co.id*, vol. 4, no. 4, 2025, doi: 10.31004/riggs.v4i4.4158.

[8] A. Kurniawan, H. S.-J. I. Sosial, and undefined 2025, "Prediksi Arah Harga Cryptocurrency Menggunakan Hybrid Lstm Encoder dan Xgboost Head: Implementasi dan Evaluasi pada 10 Aset Digital Utama," *jurnal.suriaacademicpress.com*, vol. 1, no. 2, p. 2025, 2025, Accessed: Feb. 12, 2026. [Online]. Available: https://jurnal.suriaacademicpress.com/index.php/JISEP/article/view/43

[9] D. S. Saroso, H. Rudystira, and R. Hesananda, *Manajemen Perkreditan: Berbasis Pola Data Digital Anomali Aktivitas*. Penerbit NEM, 2024.

[10] Y. Putri, A. A.-R. J. of Informatics, and undefined 2025, "Implementation of the K-Nearest Neighbor Algorithm (KNN) with Principal Component Analysis to Diagnose Tuberculosis," *journal.unnes.ac.id*, vol. 3, no. 2, 2025, doi: 10.15294/rji.v3i2.5235.

[11] P. Kegiatan, L. Di, P. Jompo, N. Umilizah, L. Iryani, and D. Fitri Brianna, "Implementation of Machine Learning Based on the K-Nearest Neighbor (KNN) Algorithm in Elderly Activities in Nursing Homes," *jurnal.polsri.ac.id*, Accessed: Feb. 12, 2026. [Online]. Available: https://jurnal.polsri.ac.id/index.php/jupiter/article/view/10592

[12] J. Al Amien, S. Sunanto, … M. R.-… (Computer S. and, and undefined 2025, "Deteksi Serangan Dalam Ekosistem Iot Melalui Analisis Multi-Class Dengan Model Xgboost Dan Penerapan Teknik Imbalance Ratio Pada Dataset IoTID20," *ejurnal.umri.ac.id*, vol. 5, no. 3, pp. 371–380, 2025, doi: 10.47065/jimat.v5i3.649.

[13] X. Berbasis *et al.*, "Algoritma Random Forest, Decision Tree, dan XGBoost untuk klasifikasi stunting pada balita," *journals.usm.ac.id*, vol. 4, no. 4, pp. 5666–5672, 2025, doi: 10.31004/riggs.v4i4.4287.

[14] M. Al-Arrafi, A. R.-J. of S. A. Social, and undefined 2025, "Implementasi Metode Algoritma Principal Component Analysis (PCA) dan Algoritma K-Nearest Neighbor (KNN) dalam Klasifikasi Buah Jambu Madu Jambu Merah," *jurnal.goretanpena.com*, no. 1, pp. 817–822, 2025, Accessed: Feb. 12, 2026. [Online]. Available: https://jurnal.goretanpena.com/index.php/JSSR/article/view/2584

[15] A. Prakoso, S. Bagus Nugroho, N. Aqiil Nugraha, F. Ferdiansyah, and I. Budiawan, "Prediksi Churn Pelanggan Telekomunikasi Menggunakan Metode Supervised Learning dengan Random Forest dan XGBoost: Penelitian," *jerkin.org*, vol. 4, pp. 17661–17667, 2026, doi: 10.31004/jerkin.v4i3.5079.

[16] A. Alkayes, T. S.-J. Pendidik. dan Teknol. Indones, and undefined 2025, "Perbandingan Algoritma XGBoost dan LSTM dalam Prediksi Harga Saham Tesla Menggunakan Data Tahun 2025," *researchgate.net*, Accessed: Feb. 12, 2026. [Online]. Available: https://www.researchgate.net/profile/Tri-Sugihartono/publication/393782620_Comparison_of_XGBoost_and_LSTM_Algorithms_in_Tesla_Stock_Price_Prediction_Using_2025_Data/links/6879a6864f72461c714f2c17/Comparison-of-XGBoost-and-LSTM-Algorithms-in-Tesla-Stock-Price-Prediction-Using-2025-Data.pdf

[17] S. Frastio, … M. N.-S. N. M., and undefined 2025, "Penerapan Algoritma Xgboost Untuk Analisis Dan Klasifikasi Risiko Kredit Berbasis Data Mining," *senadimu.potensi-utama.org*, Accessed: Feb. 12, 2026. [Online]. Available: https://senadimu.potensi-utama.org/index.php/home/article/view/561

[18] F. Hidayat, … N. S.-R. J. of, and undefined 2026, "Perbandingan Algoritma XGBoost dan Random Forest dalam Klasifikasi Surat Masuk Pemerintahan," *journal.ilmudata.co.id*, Accessed: Feb. 12, 2026. [Online]. Available: https://journal.ilmudata.co.id/index.php/RIGGS/article/view/5044

[19] M. Nururrohman, E. S.-S. J. Informatika, and undefined 2025, "Sistem Rekomendasi Penerima Beasiswa pada Calon Mahasiswa Menggunakan Algoritma XGBoost di Universitas Islam Darul 'Ulum Lamongan," *journal-cs.unsika.ac.id*, Accessed: Feb. 12, 2026. [Online]. Available: https://journal-cs.unsika.ac.id/index.php/syntax/article/view/13-24

[20] R. RAHMAN, "IMPLEMENTASI METODE STACKING MENGGUNAKAN RANDOM FOREST DAN XGBOOST UNTUK MENGKLASIFIKASI RISIKO SERANGAN JANTUNG," 2025, Accessed: Feb. 12, 2026. [Online]. Available: https://repository.mercubuana.ac.id/96574/