



## ***Development of a Classification Model to Evaluate Strategic Laundry Business Locations Using Geospatial and Open-Source Data***

### **Pembangunan Model Klasifikasi untuk Menilai Lokasi Strategis Usaha Laundry Berdasarkan Analisis Geospasial dan Data Open-Source**

Briandana Riznov<sup>1\*</sup>, Bagus Jati Santoso<sup>2</sup>

<sup>1,2</sup>Program Magister Manajemen Teknologi, Sekolah Interdisiplin Manajemen dan Teknologi, Institut Teknologi Sepuluh Nopember (ITS), Indonesia

E-Mail: <sup>1</sup>6032231214@student.its.ac.id, <sup>2</sup>bagus@if.its.ac.id

Received Nov 09th 2025; Revised Dec 12th 2025; Accepted Dec 28th 2025; Available Online Jan 15th 2026

Corresponding Author: Briandana Riznov

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

#### **Abstract**

The laundry service industry in Indonesia continues to grow rapidly in line with the increasingly practical lifestyle of urban communities. However, many entrepreneurs still face business failures due to inaccurate location selection, highlighting the need for a data-driven evaluation framework. This study aims to develop a classification model based on geospatial and open-source data to assess whether existing laundry business locations in Surabaya are categorized as strategic or non-strategic. The research collected open data from Google Maps API, OpenStreetMap, and the Central Bureau of Statistics, including spatial, demographic, and economic indicators. The analytical process involved data preprocessing, feature engineering, data balancing using SMOTE, and the application of five classification algorithms—Logistic Regression, Decision Tree, Naive Bayes, k-Nearest Neighbor, and Random Forest—evaluated through metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC. The Random Forest algorithm achieved the best performance with an accuracy of 90.1% and ROC-AUC of 0.957. The findings indicate that population density, proximity to residential areas, and business competition significantly influence location feasibility. This research contributes to building a geospatial and data-based evaluation framework to support objective, evidence-based decision-making in the service business sector, particularly for urban micro-enterprises in Indonesia.

**Keyword:** Classification Model, Geospatial Analysis, Open-Source Data, Random Forest, Strategic Business

#### **Abstrak**

Industri jasa cuci pakaian di Indonesia terus mengalami pertumbuhan pesat seiring perubahan gaya hidup masyarakat urban yang semakin praktis. Namun, banyak pelaku usaha mengalami kegagalan akibat pemilihan lokasi yang kurang tepat, sehingga diperlukan pendekatan berbasis data untuk mengevaluasi kelayakan lokasi usaha secara objektif. Penelitian ini bertujuan untuk mengembangkan model klasifikasi berbasis analisis geospasial dan data open-source guna menilai apakah lokasi usaha *laundry eksisting* di Surabaya tergolong strategis atau tidak strategis. Data penelitian dikumpulkan dari Google Maps API, OpenStreetMap (OSM), dan Badan Pusat Statistik (BPS), yang mencakup indikator spasial, demografis, dan ekonomi. Proses analisis meliputi pembersihan data, rekayasa fitur, penyeimbangan data dengan SMOTE, serta penerapan lima algoritma klasifikasi Logistic Regression, Decision Tree, Naive Bayes, k-Nearest Neighbor, dan Random Forest dengan evaluasi menggunakan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan akurasi 90,1% dan ROC-AUC 0,957. Faktor kepadatan penduduk, kedekatan dengan permukiman, dan jumlah kompetitor terbukti paling berpengaruh terhadap tingkat strategis lokasi. Penelitian ini berkontribusi dalam menyediakan kerangka evaluasi berbasis data dan geospasial yang dapat digunakan sebagai dasar pengambilan keputusan bisnis yang lebih objektif, terukur, dan adaptif terhadap dinamika perkotaan.

Kata Kunci: Analisis Geospasial, , Bisnis Strategis, Data Open-Source, Model Klasifikasi, Random Forest

#### **1. PENDAHULUAN**

Industri jasa cuci pakaian di Indonesia menunjukkan pertumbuhan yang pesat dalam beberapa tahun terakhir seiring dengan perubahan gaya hidup masyarakat perkotaan yang semakin praktis dan efisien. Di



kota besar seperti Surabaya, kebutuhan terhadap layanan *laundry* yang cepat, terjangkau, dan mudah diakses telah menjadi bagian dari aktivitas kehidupan sehari-hari masyarakat urban. Bisnis jasa cuci pakaian kiloan juga dikenal sebagai salah satu sektor jasa rumahan yang memiliki prospek menjanjikan, terutama di kalangan mahasiswa, keluarga muda, dan pekerja kantoran yang memiliki keterbatasan waktu untuk mengurus pekerjaan domestik. Secara makro, industri perawatan cucian di Indonesia diperkirakan memiliki nilai ekonomi yang besar dengan tingkat pertumbuhan yang stabil dalam beberapa tahun ke depan, di mana segmen jasa *laundry* menjadi salah satu kontributor utama. Kondisi ini menunjukkan bahwa sektor jasa *laundry* memiliki potensi ekonomi yang signifikan dalam mendukung pertumbuhan industri jasa rumah tangga di wilayah perkotaan.

Meskipun memiliki prospek yang menjanjikan, banyak pelaku usaha *laundry* masih mengalami kesulitan dalam mempertahankan keberlanjutan usaha akibat tidak tepatnya dalam pemilihan lokasi. Dalam konteks perkotaan yang padat seperti Surabaya, keberhasilan bisnis tidak hanya ditentukan oleh harga dan kualitas layanan, tetapi juga oleh strategi penentuan lokasi yang tepat. Lokasi yang strategis ditandai dengan tingkat aksesibilitas yang baik, kedekatan dengan area permukiman padat, serta tingkat persaingan yang moderat. Sebaliknya, lokasi yang kurang tepat dapat menyebabkan rendahnya permintaan pelanggan, tingginya biaya operasional, hingga berujung pada kegagalan usaha. Oleh karena itu, pemahaman terhadap karakteristik spasial dan demografis wilayah sekitar menjadi aspek penting dalam pengambilan keputusan berbasis data.

Berbagai penelitian terdahulu telah mengkaji pemilihan lokasi bisnis menggunakan pendekatan *geospasial* dan *machine learning*. Penelitian [1] menunjukkan bahwa integrasi variabel spasial dan demografis dapat meningkatkan kemampuan model dalam memahami dinamika kawasan urban. Penelitian oleh *Zheng* [2] menggunakan regresi logistik berbasis GIS untuk menilai kelayakan lokasi fasilitas publik seperti *parcel-pickup lockers* dan menemukan bahwa faktor jarak, aksesibilitas, dan kepadatan penduduk memengaruhi akurasi prediksi spasial. Dalam konteks ritel, *Zhao* [3] memanfaatkan data *multi-sumber* dan algoritma *machine learning* untuk memprediksi lokasi optimal *coffee shop*, sedangkan *Gao* [4] membandingkan pola persebaran gerai antara *Starbucks* dan *Luckin Coffee* dan menunjukkan bahwa preferensi lokasi bisnis dapat dipahami melalui analisis spasial berbasis data.

Selain itu, pendekatan *hybrid* dalam pemodelan performa ritel juga telah banyak dikaji. *Lagzi* [5] mengembangkan model prediksi performa usaha dengan menggabungkan metode *stochastic DEA* dan *decision tree*, sementara *Lu et al* [6] menerapkan model *screening* berbasis *machine learning* untuk mengidentifikasi kelayakan lokasi toko secara lebih efisien. Pendekatan *ensemble* seperti yang digunakan *Samal & Ghosh* [7] juga menunjukkan efektivitasnya dalam meningkatkan akurasi prediksi permintaan ritel pada lingkungan *multi-channel*. Meskipun penelitian-penelitian tersebut memberikan kontribusi signifikan dalam pemahaman spasial dan pemodelan keputusan bisnis, sebagian besar masih berfokus pada usaha berskala besar dan belum secara spesifik diterapkan pada sektor usaha mikro seperti *laundry*.

Selain itu, perkembangan teknologi pemetaan digital dan data *open-source* membuka peluang besar untuk analisis spasial berbasis data secara lebih terjangkau. *Chen et al* [8] menegaskan bahwa integrasi data POI, kepadatan penduduk, mobilitas, serta aktivitas wilayah mampu memberikan gambaran yang lebih akurat terhadap potensi pasar suatu area. Di sisi lain, model *deep learning* dan *ensemble* seperti *Random Forest* terbukti berhasil dalam mengidentifikasi *hotspot* komersial, memetakan zona fungsional perkotaan, serta memprediksi performa lokasi berdasarkan karakteristik lingkungan. Namun demikian, pemanfaatan pendekatan tersebut dalam konteks UMKM seperti *laundry* masih relatif terbatas, terutama di Indonesia.

Dari tinjauan penelitian tersebut, dapat diidentifikasi adanya *research gap*, yaitu belum tersedianya model evaluasi lokasi usaha *laundry* berbasis analisis *geospasial* yang mengintegrasikan variabel spasial, demografis, dan ekonomi secara komprehensif dengan dukungan model klasifikasi terawasi. Penelitian ini hadir untuk mengisi kekosongan tersebut dengan mengembangkan model klasifikasi berbasis data terbuka yang mampu menilai tingkat strategis lokasi *laundry* di Surabaya secara kuantitatif. Selain itu, pendekatan visualisasi seperti peta interaktif, *heatmap*, dan *dashboard* analitik digunakan untuk mempermudah interpretasi hasil dan mendukung proses pengambilan keputusan bagi pelaku usaha maupun pemangku kepentingan.

Secara keseluruhan, penelitian ini berkontribusi dalam pengembangan kerangka evaluasi berbasis data dan analisis *geospasial* yang dapat digunakan untuk menilai kelayakan lokasi usaha jasa di lingkungan perkotaan. Hasil penelitian diharapkan menjadi referensi bagi pelaku usaha *laundry* maupun pemerintah daerah dalam merumuskan kebijakan tata ruang bisnis yang lebih efisien, adaptif, dan berkelanjutan.

## 2. METODOLOGI PENELITIAN

Metodologi penelitian merupakan fondasi penting yang menentukan ketepatan hasil dan kualitas interpretasi dalam suatu kajian ilmiah. Pada penelitian ini, pendekatan yang digunakan disusun secara sistematis untuk memastikan bahwa seluruh proses analisis, mulai dari pemahaman masalah hingga pengujian model, dilakukan secara konsisten, terukur, dan dapat *direplikasi*. Dengan mengintegrasikan *machine learning*, analisis *geospasial*, serta pemanfaatan data terbuka, penelitian ini mengadopsi metodologi

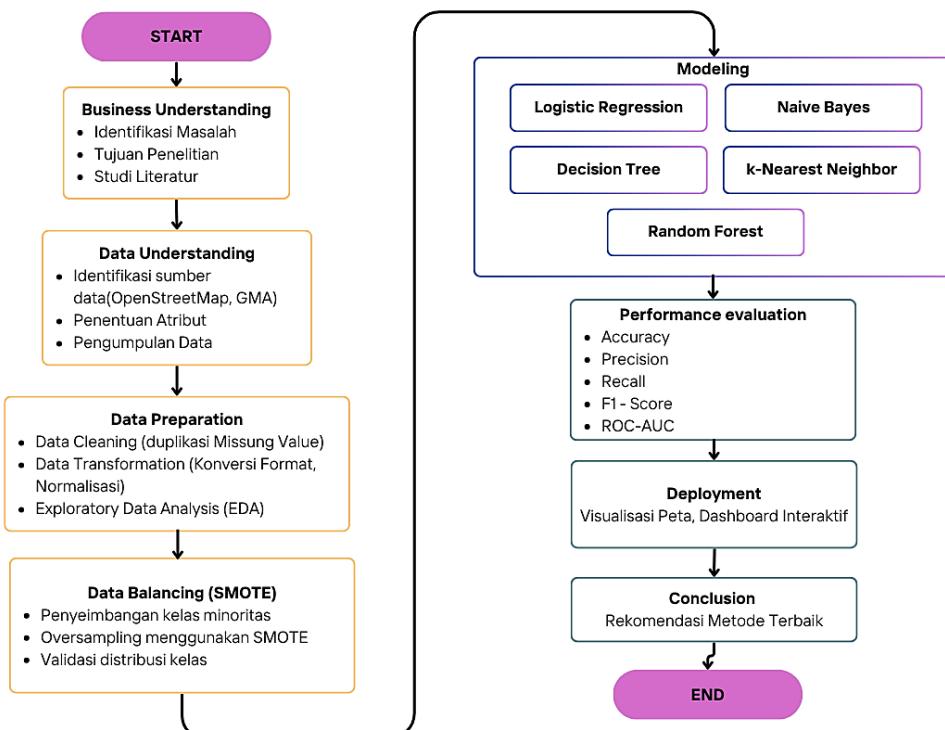
yang tidak hanya relevan dengan perkembangan ilmiah terkini, tetapi juga mampu menjawab tantangan nyata dalam pemilihan lokasi usaha berbasis data.

Pendekatan metodologis ini dirancang untuk membangun kerangka evaluasi yang komprehensif, di mana setiap tahapan berperan dalam menghasilkan pemodelan klasifikasi yang reliabel. Tahap-tahap tersebut tidak hanya menekankan aspek teknis, tetapi juga mempertimbangkan karakteristik spasial dan sosial ekonomi yang memengaruhi kelayakan lokasi *laundry* di wilayah perkotaan. Dengan demikian, metodologi penelitian ini menghadirkan kombinasi antara ketelitian ilmiah dan penerapan praktis, sehingga mampu memberikan kontribusi yang signifikan bagi pengembangan *data-driven decision making* dalam konteks bisnis dan tata ruang.

## 2.1. Kerangka Penelitian

Penelitian ini menggunakan pendekatan analisis *geospasial* berbasis data terbuka (*open-source*) untuk menilai tingkat strategis lokasi usaha *laundry eksisting* di Surabaya. Pendekatan ini sejalan dengan perkembangan riset *geospasial* yang menekankan pemanfaatan data *multi-sumber* dan metodologi *machine learning* dalam memahami pola spasial usaha komersial [1][2][3]. Untuk memastikan proses analisis berjalan secara sistematis dan terstruktur, penelitian ini mengadopsi kerangka kerja CRISP-DM yang terdiri atas enam tahap utama, yaitu Business *Understanding*, Data *Understanding*, Data *Preparation*, Modeling, Evaluation, dan Deployment. Pada penelitian ini, tahap Data *Preparation* diperluas dengan proses penyeimbangan data (*Data Balancing*) untuk menangani ketidakseimbangan kelas, serta diakhiri dengan penyusunan kesimpulan penelitian [4][5].

Setiap tahap dalam CRISP-DM dirancang untuk memastikan bahwa data *geospasial* dan demografis yang digunakan telah diproses secara tepat sehingga model klasifikasi dapat bekerja secara optimal. Pada tahap Data *Preparation*, misalnya, data lokasi *laundry*, jarak ke permukiman, kepadatan penduduk, intensitas kompetitor, dan variabel lingkungan lainnya diolah menjadi fitur spasial terstruktur sesuai dengan praktik terbaik dalam pemodelan lokasi bisnis [6]. Sementara itu, tahap Data *Balancing* menerapkan teknik *oversampling* seperti SMOTE untuk mengatasi ketidakseimbangan kelas, sebagaimana direkomendasikan dalam penelitian-penelitian terbaru mengenai klasifikasi spasial [7].



Gambar 1. Alur metodologi penelitian

Gambar 1 tidak hanya menggambarkan urutan proses, tetapi juga menjelaskan keterkaitan antar tahapan, sehingga memperkuat pemahaman mengenai bagaimana data spasial diproses hingga menghasilkan model klasifikasi lokasi strategis. Pada Gambar 1, proses Data *Balancing* ditampilkan secara terpisah untuk memperjelas tahapan teknis penanganan ketidakseimbangan kelas, meskipun secara konseptual merupakan bagian dari tahap Data *Preparation* dalam kerangka CRISP-DM.

Tahapan-tahapan tersebut dirancang untuk membangun model klasifikasi yang mampu menilai apakah suatu lokasi *laundry* tergolong strategis atau tidak strategis berdasarkan variabel spasial, demografis, dan ekonomi yang diperoleh dari platform data *open-source* seperti Google Maps API, OpenStreetMap (OSM), dan Badan Pusat Statistik (BPS). Integrasi data ini sejalan dengan penelitian terdahulu yang menunjukkan bahwa kombinasi faktor lingkungan dan sosial-ekonomi dapat meningkatkan akurasi prediksi lokasi usaha [3][4]. Dengan demikian, kerangka penelitian ini tidak hanya menyediakan dasar metodologis yang kuat, tetapi juga memastikan bahwa model yang dibangun dapat memberikan rekomendasi yang aplikatif bagi pelaku usaha dan pemangku kebijakan.

Penelitian mengenai pemilihan lokasi usaha berbasis data spasial telah banyak dikembangkan dengan memanfaatkan pendekatan *machine learning* maupun analisis *geospasial*. Georgati *et al* [1] menunjukkan bahwa penggunaan variabel spasial dan demografis secara terpadu dapat meningkatkan kemampuan model dalam memetakan dinamika wilayah perkotaan. Sementara itu, Zheng *et al* [2] menggunakan pendekatan regresi logistik berbasis GIS untuk menentukan kelayakan lokasi fasilitas publik dan membuktikan bahwa jarak, aksesibilitas, serta kepadatan penduduk merupakan variabel penting dalam pemodelan spasial.

Dalam konteks bisnis ritel, [3] menerapkan *machine learning* dengan data *multi-sumber* untuk memprediksi lokasi optimal *coffee shop*, sedangkan [4] membandingkan sebaran lokasi *Starbucks* dan *Luckin Coffee* menggunakan teknik pemodelan spasial berbasis data besar. Penelitian ini menggaris bawahi pentingnya integrasi variabel lingkungan serta pola aktivitas wilayah dalam memahami preferensi lokasi bisnis modern.

*Logistic Regression* banyak digunakan dalam studi prediktif yang membutuhkan interpretasi koefisien model, sementara *Decision Tree* dan *Random Forest* terbukti efektif dalam menangani hubungan non-linier serta menghasilkan informasi kepentingan fitur pada konteks pemodelan lokasi usaha [6]. Pendekatan *Naive Bayes* dan k-NN juga telah digunakan dalam berbagai aplikasi klasifikasi spasial dan ritel berbasis data *multi-sumber* karena efisiensi komputasinya [7]. Penelitian oleh Lu *et al* [9] juga menegaskan efektivitas model *ensemble* dalam meningkatkan akurasi prediksi berbasis data *multi-channel*.

Salah satu teknik yang umum digunakan adalah *Synthetic Minority Oversampling Technique* (SMOTE), yang telah banyak diterapkan dalam studi klasifikasi spasial dan ritel untuk mengatasi ketidakseimbangan kelas [10]. Secara keseluruhan, *literature review* ini menunjukkan bahwa meskipun pemilihan lokasi dengan analisis *geospasial* dan *machine learning* telah banyak dikembangkan, penerapannya masih terbatas pada sektor ritel berskala besar. Dengan demikian, penelitian ini hadir untuk mengisi celah tersebut dengan fokus pada usaha mikro seperti *laundry*, berbasis data *open-source* yang lebih mudah diakses dan *direplikasi*.

## 2.2. Business Understanding

Tahap Business *Understanding* bertujuan untuk mengidentifikasi permasalahan utama, kebutuhan pengguna, serta ruang lingkup analisis yang akan dilakukan. Permasalahan inti dalam penelitian ini adalah belum tersedianya model evaluasi lokasi bisnis *laundry* yang bersifat objektif, terukur, serta mampu mengintegrasikan variabel spasial, demografis, dan ekonomi sebagai dasar pengambilan keputusan. Dalam konteks kota besar seperti Surabaya, pemilihan lokasi usaha *laundry* sering kali masih bersifat intuitif dan tidak berbasis data, sehingga berpotensi menimbulkan tidak *efisien* dan risiko kegagalan usaha.

Penelitian terdahulu menunjukkan bahwa faktor spasial seperti kepadatan penduduk, aksesibilitas jaringan jalan, intensitas kompetitor, serta pola aktivitas komersial merupakan determinan penting dalam keberhasilan suatu usaha jasa [1][2]. Berbagai kajian berbasis *machine learning* dan *geospasial* juga membuktikan bahwa pemodelan berbasis data *multi-sumber* mampu meningkatkan akurasi prediksi lokasi bisnis dan mengidentifikasi area potensial secara lebih komprehensif [3][4]. Sebagai contoh, Zhao *et al.* [3] mengembangkan model prediksi lokasi *coffee shop* berbasis data *multi-sumber*, sementara Lagzi *et al.* [5] menunjukkan bagaimana analisis pola lokasi ritel dapat digunakan untuk memahami preferensi konsumen pada skala perkotaan. Selain itu, penelitian oleh Geo [4] dan Zhao [3] memperkuat pandangan bahwa integrasi variabel spasial dan demografis penting dalam memahami karakteristik lingkungan bisnis.

Berdasarkan kebutuhan penelitian dan temuan ilmiah dari berbagai studi terdahulu, penelitian ini bertujuan untuk menganalisis berbagai faktor spasial, demografis, dan ekonomi yang berpengaruh terhadap tingkat strategis lokasi usaha *laundry* di Surabaya. Analisis ini mencakup karakteristik lingkungan seperti kepadatan penduduk, aksesibilitas jaringan jalan, persebaran kompetitor, serta kondisi sosial-ekonomi yang menjadi determinan penting dalam keberhasilan usaha berbasis lokasi. Selain itu, penelitian ini juga bertujuan mengembangkan model klasifikasi berbasis data spasial untuk menilai tingkat strategis lokasi usaha *laundry* dengan memanfaatkan berbagai pendekatan klasifikasi yang umum digunakan dalam pemodelan lokasi bisnis. Pendekatan ini dipilih karena mampu menangani karakteristik data spasial yang bersifat linier maupun non-linier serta telah banyak diterapkan dalam studi pemilihan lokasi usaha [5][6][7]. Tujuan selanjutnya adalah mengevaluasi performa setiap model menggunakan metrik akurasi, presisi, *recall*, dan F1-score, sehingga dapat diidentifikasi model terbaik yang mampu memberikan hasil prediksi paling reliabel.

Dengan demikian, penelitian ini diharapkan dapat menghasilkan sistem evaluasi lokasi usaha yang lebih objektif, adaptif, serta mendukung proses pengambilan keputusan berbasis data.

Selain itu, tahapan ini juga mencakup pemahaman kebutuhan pengguna, yaitu pelaku usaha *laundry*, pemerintah daerah, dan calon investor yang membutuhkan sistem pendukung keputusan berbasis data spasial. Mengacu pada penelitian [9] penggunaan model prediktif tidak hanya membantu memahami pola permintaan tetapi juga meningkatkan keandalan keputusan bisnis di sektor jasa. Penelitian [10] serta [11]. menunjukkan bahwa model klasifikasi spasial mampu memetakan zona komersial secara presisi, sehingga dapat diadaptasikan dalam konteks penilaian lokasi usaha mikro seperti *laundry*.

Dengan demikian, tahap Business *Understanding* memastikan bahwa seluruh proses penelitian diarahkan untuk mengatasi kebutuhan nyata di lapangan serta mengisi *research gap* yang diidentifikasi, yaitu terbatasnya kajian pemilihan lokasi usaha mikro berbasis data *open-source* dengan pendekatan klasifikasi terawasi pada skala perkotaan di Indonesia.

### **2.3. Data Understanding**

Informasi mengenai lokasi usaha *laundry* serta persebaran kompetitor diperoleh melalui Google *Maps API*, yang digunakan untuk mengidentifikasi titik usaha *eksisting* dan atribut spasial berbasis *point-of-interest* (POI) [1][11]. Sementara itu, data jaringan jalan, struktur wilayah, serta batas administratif diperoleh dari OSM, sebuah platform pemetaan kolaboratif yang sering dimanfaatkan untuk analisis aksesibilitas dan pemodelan spasial karena kelengkapan dan keterbukaan datanya [2][12].

Untuk melengkapi aspek demografis dan sosial ekonomi, penelitian ini menggunakan data BPS seperti jumlah penduduk, kepadatan penduduk, jumlah rumah tangga, serta karakteristik wilayah lainnya yang relevan dalam menentukan potensi permintaan layanan *laundry*. Variabel-variabel ini telah terbukti berperan penting dalam prediksi lokasi usaha dan perilaku pasar berdasarkan penelitian terdahulu yang memanfaatkan indikator demografis dalam konteks klasifikasi spasial [5]. Selain itu, data harga sewa rumah dan kos diperoleh dari platform OLX Indonesia sebagai indikator proksi kondisi ekonomi lokal, khususnya untuk menggambarkan variasi biaya hunian antarwilayah. Pendekatan penggunaan data daring sebagai proksi ekonomi lokal juga telah digunakan dalam berbagai studi penilaian kelayakan lokasi usaha berbasis parameter biaya dan karakteristik hunian.

Secara keseluruhan, penelitian ini memanfaatkan 801 titik lokasi laundry yang tersebar di 10 kecamatan utama di Kota Surabaya, yang merepresentasikan wilayah dengan konsentrasi usaha laundry dan aktivitas permukiman yang tinggi. Pemilihan kombinasi variabel ini mengikuti pendekatan yang dilakukan dalam penelitian-penelitian sebelumnya yang mengintegrasikan POI, demografi, dan struktur ruang kota untuk meningkatkan akurasi model prediktif lokasi usaha [13]. Dengan memahami karakteristik data sejak tahap awal, penelitian ini memastikan bahwa proses pemodelan selanjutnya dapat dilakukan secara tepat, terarah, dan menghasilkan prediksi lokasi strategis yang dapat diandalkan.

### **2.4. Data Preparation**

Tahap Data *Preparation* dilakukan untuk memastikan bahwa seluruh variabel yang digunakan dalam penelitian telah siap dan layak untuk dianalisis dalam proses pemodelan. Proses ini dimulai dengan pembersihan data (*data cleaning*), yaitu menghilangkan data duplikat, mendeteksi tidak *konsisten*, serta menangani nilai hilang (*missing values*) menggunakan pendekatan *imputasi* median untuk variabel numerik dan modus untuk variabel *kategorikal*. Pendekatan ini banyak digunakan dalam penelitian klasifikasi spasial karena relatif robust terhadap *outlier* dan dapat menjaga distribusi data secara umum tanpa mengubah struktur utama *dataset* [4].

Selanjutnya dilakukan proses transformasi data yang meliputi normalisasi variabel numerik serta *pengkodean* (*encoding*) variabel *kategorikal* menggunakan metode label *encoding*. Normalisasi diperlukan agar algoritma seperti k-NN dan *Logistic Regression* dapat bekerja lebih stabil, mengingat beberapa model sensitif terhadap skala fitur. Sementara itu, label *encoding* membantu mengubah kategori non-numerik menjadi nilai numerik sehingga dapat diproses oleh algoritma klasifikasi seperti *Decision Tree* dan *Random Forest* [5].

Tahap berikutnya adalah *Exploratory Data Analysis* (EDA), yang mencakup analisis distribusi data, identifikasi *outlier*, serta pengamatan korelasi *antarvariabel*. EDA dilakukan untuk memahami pola spasial dan hubungan *antarfitur* sebelum data masuk ke tahap modeling. Dalam penelitian-penelitian sebelumnya, EDA terbukti menjadi langkah penting untuk mengidentifikasi fitur yang paling berkontribusi dalam pemodelan lokasi usaha dan zona komersial [7][10]. Melalui proses ini, peneliti dapat menilai kelayakan setiap variabel, mendeteksi anomali, serta memastikan bahwa struktur data telah sesuai dengan karakteristik metode klasifikasi yang digunakan.

Secara keseluruhan, tahap Data *Preparation* memastikan bahwa *dataset* yang berisi 801 titik lokasi *laundry* dengan 22 atribut spasial, demografis, dan ekonomi berada dalam kondisi optimal untuk diproses lebih lanjut pada tahap Data *Balancing* dan *Modeling*. Tahapan ini mengikuti praktik terbaik yang digunakan

dalam berbagai penelitian *geospasial* dan pemodelan lokasi berbasis *machine learning* [7], sehingga mampu menghasilkan model klasifikasi yang lebih stabil dan akurat.

### 2.5. Penyeimbangan Data (SMOTE)

*Dataset* yang digunakan dalam penelitian ini menunjukkan distribusi kelas yang tidak seimbang antara kategori lokasi strategis dan tidak strategis. Ketidakseimbangan ini berpotensi menghasilkan bias model terhadap kelas mayoritas, sehingga diperlukan teknik penyeimbangan kelas sebelum proses pemodelan dilakukan. Untuk mengatasi masalah tersebut, penelitian ini menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*), yaitu teknik yang menghasilkan sampel sintetis berdasarkan titik terdekat dari kelas minoritas melalui pendekatan *k-nearest neighbors*. Tidak seperti metode *oversampling* tradisional yang hanya melakukan duplikasi data, SMOTE menciptakan data sintetis yang lebih representatif sehingga berpotensi membantu model mengenali pola kelas minoritas dengan lebih baik [8].

Penerapan SMOTE telah banyak digunakan dalam penelitian pemodelan ritel dan klasifikasi spasial [9], misalnya, Ketidakseimbangan kelas juga banyak dilaporkan dalam penelitian pemodelan ritel dan klasifikasi spasial, terutama ketika distribusi aktivitas ekonomi tidak merata secara geografis [13][14]. Kondisi ini dapat menurunkan performa algoritma prediksi apabila tidak ditangani secara tepat, sehingga diperlukan strategi penanganan *imbalance* data dalam proses pemodelan lokasi usaha. Selain itu, penelitian oleh Wang, J. & Huang [15] pada pemetaan *hotspot* komersial juga menunjukkan bahwa ketidakseimbangan frekuensi lokasi dapat menurunkan performa algoritma prediksi, sehingga proses *balancing* diperlukan untuk meningkatkan kualitas model.

Dalam konteks pemilihan lokasi usaha *laundry*, distribusi data yang tidak merata antara area strategis dan tidak strategis merupakan fenomena yang umum ditemukan di wilayah perkotaan. Beberapa lokasi cenderung memiliki konsentrasi usaha yang tinggi, sementara lokasi lain minim aktivitas ekonomi. Dengan menerapkan SMOTE pada data latih dalam setiap *fold cross-validation*, penelitian ini memastikan bahwa model klasifikasi tidak hanya belajar dari pola mayoritas, tetapi juga mampu mengenali karakteristik lokasi minoritas tanpa menimbulkan data *leakage* antar data latih dan uji. Pendekatan ini selaras dengan praktik terbaik yang digunakan dalam pemodelan lokasi komersial dan layanan perkotaan berbasis *machine learning* [14].

### 2.6. Pemodelan

Tahap pemodelan dilakukan menggunakan lima algoritma klasifikasi yang banyak digunakan dalam penelitian analitik spasial dan pemodelan lokasi berbasis *machine learning*. Model pertama adalah *Logistic Regression*, yang umum digunakan dalam berbagai studi prediksi berbasis data spasial karena kemampuannya dalam memberikan interpretasi yang jelas terhadap pengaruh masing-masing variabel. Model ini relevan dalam konteks pemilihan lokasi usaha untuk mengevaluasi hubungan linier antara variabel spasial dan tingkat strategis lokasi. Selanjutnya, *Decision Tree* digunakan karena kemampuannya dalam membentuk aturan keputusan yang mudah dipahami serta menangani variabel numerik maupun *kategorikal* secara fleksibel [9]. Pendekatan *Naive Bayes* juga digunakan sebagai model pembanding yang efisien dan bersifat *probabilistik*, terutama untuk dataset dengan jumlah fitur yang relatif besar [7].

Algoritma *k-Nearest Neighbor* (*k-NN*) digunakan sebagai pendekatan berbasis kedekatan pada ruang fitur untuk menangkap pola kemiripan antar lokasi. Model ini sering digunakan sebagai pembanding dalam studi klasifikasi spasial karena kesederhanaan dan sensitivitasnya terhadap struktur data. Terakhir, penelitian ini menggunakan *Random Forest*, salah satu metode *ensemble learning* yang terdiri dari banyak *Decision Tree* dan dikenal memiliki stabilitas prediksi yang tinggi, kemampuan menangani hubungan non-linier, serta menghasilkan informasi *feature importance* yang bermanfaat dalam memahami faktor-faktor penentu lokasi strategis [11]. Model *Random Forest* telah banyak digunakan dalam analisis spasial dan pemodelan lokasi usaha karena performanya yang konsisten pada data heterogen.

Untuk menjamin reliabilitas hasil dan menghindari bias akibat pembagian data yang tidak konsisten, Seluruh algoritma diuji menggunakan metode *5-Fold Cross Validation*, di mana proses pelatihan dan penyeimbangan data dilakukan secara terpisah pada setiap *fold* untuk menghindari data *leakage* dan memastikan estimasi performa yang lebih stabil dan representatif [16]. Dengan demikian, penggunaan kelima algoritma tersebut diharapkan dapat memberikan gambaran yang komprehensif mengenai kemampuan model dalam menilai tingkat strategis lokasi usaha *laundry* berbasis data spasial, demografis, dan ekonomi.

### 2.7. Evaluasi Kinerja

Evaluasi kinerja model dilakukan menggunakan beberapa metrik utama, yaitu *Accuracy*, *Precision*, *Recall*, F1-Score, dan ROC-AUC, yang secara luas digunakan dalam penelitian klasifikasi biner dan pemodelan spasial berbasis *machine learning*. *Accuracy* digunakan untuk mengukur proporsi prediksi yang benar secara keseluruhan, sementara *Precision* dan *Recall* mencerminkan kemampuan model dalam mengidentifikasi lokasi yang diklasifikasikan sebagai strategis. F1-Score digunakan sebagai metrik keseimbangan antara *Precision* dan *Recall*, terutama pada kondisi *dataset* yang tidak seimbang. ROC-AUC

digunakan untuk menilai kemampuan model dalam membedakan kelas strategis dan tidak strategis pada berbagai *threshold* prediksi.

Setiap model dievaluasi menggunakan 5-Fold *Cross Validation* untuk memastikan reliabilitas metrik yang dihasilkan serta menghindari bias akibat pembagian data yang tidak konsisten [10]. Pendekatan ini terbukti efektif dalam penelitian *geospasial* dan prediksi lokasi ritel, di mana distribusi spasial sering kali tidak merata [13]. Hasil evaluasi menunjukkan bahwa *Random Forest* memberikan performa paling konsisten dengan nilai *Accuracy*, F1-Score, dan ROC-AUC yang relatif lebih tinggi dibandingkan model lainnya, sehingga dipilih sebagai model dengan kinerja terbaik pada penelitian ini. Temuan ini sejalan dengan penelitian terdahulu yang melaporkan stabilitas dan keandalan *Random Forest* dalam menangani hubungan non-linier pada data spasial [16].

### 2.8. Deployment dan Visualisasi

Tahap *deployment* dalam penelitian ini dilakukan dalam bentuk penyajian hasil klasifikasi model melalui platform visualisasi berbasis *geospasial*, yang menampilkan persebaran lokasi strategis dan tidak strategis di wilayah Surabaya. Visualisasi ini berupa peta interaktif, *heatmap*, dan *dashboard* analitik yang memungkinkan pengguna melakukan eksplorasi spasial serta mengamati pola konsentrasi usaha. Pendekatan visual ini sejalan dengan penelitian *geospasial* terdahulu yang menekankan pentingnya representasi spasial sebagai alat bantu dalam proses pengambilan keputusan.

Peta interaktif disusun menggunakan data koordinat lokasi *laundry*, atribut demografis, serta nilai prediksi model. *Heatmap* digunakan untuk menunjukkan area dengan tingkat kepadatan usaha serta indikasi potensi permintaan berdasarkan hasil klasifikasi model dan karakteristik lingkungan sekitarnya, sedangkan *dashboard* analitik menampilkan statistik pendukung seperti jumlah kompetitor, tingkat kepadatan penduduk, dan nilai probabilitas lokasi strategis [17], serta memungkinkan pemangku kepentingan untuk memahami dinamika spasial secara intuitif dan informatif. Dengan demikian, hasil model tidak hanya berhenti pada analisis numerik, tetapi juga dapat diimplementasikan secara praktis dalam konteks perencanaan usaha dan kebijakan tata ruang kota.

### 2.9. Kesimpulan Metodologi

Metodologi penelitian ini menggabungkan pendekatan *machine learning*, analisis *geospasial*, dan visualisasi data untuk menghasilkan kerangka evaluasi lokasi usaha yang objektif, komprehensif, dan dapat *direplikasi*. Dengan mengintegrasikan variabel spasial, demografis, dan ekonomi, penelitian ini menawarkan pendekatan berbasis data yang lebih adaptif dibandingkan metode tradisional yang bersifat intuitif. Proses CRISP-DM digunakan sebagai kerangka utama untuk memastikan bahwa setiap tahap penelitian, mulai dari *Business Understanding* hingga *Deployment*, berjalan secara sistematis dan dapat ditelusuri kembali.

Penerapan lima algoritma klasifikasi memberikan perspektif komparatif terhadap performa model dalam konteks data spasial. Evaluasi model melalui 5-Fold *Cross Validation* memastikan bahwa hasil yang diperoleh bersifat konsisten dan tidak bias, sementara penggunaan SMOTE membantu mengatasi ketidakseimbangan data yang sering muncul dalam studi spasial perkotaan. Integrasi visualisasi berbasis *geospasial* memungkinkan hasil model disajikan secara intuitif dan mudah dipahami oleh pelaku usaha, pemerintah daerah, maupun pihak lain yang membutuhkan analisis spasial sebagai bahan pertimbangan dalam pengambilan keputusan. Metodologi ini sejalan dengan penelitian terdahulu mengenai pemetaan zona komersial, prediksi lokasi ritel, dan analisis dinamika spasial [18][19], sehingga memperkuat kontribusi metodologi penelitian ini dalam pengembangan evaluasi lokasi usaha berbasis data di wilayah perkotaan.

## 3. HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan merupakan tahap penting dalam penelitian berbasis pemodelan klasifikasi, karena pada tahap ini kinerja model, pola data, serta implikasi hasil analisis dievaluasi secara sistematis. Dalam konteks analisis spasial dan pengambilan keputusan berbasis data, penyajian hasil tidak hanya berfokus pada nilai metrik evaluasi, tetapi juga pada *keterjelasan* interpretasi model dan keterkaitannya dengan fenomena spasial yang diamati. *Klimonová et al.* [20] menekankan bahwa penyajian hasil yang terstruktur dan didukung visualisasi spasial berperan penting dalam meningkatkan *interpretabilitas* model klasifikasi, khususnya pada pendekatan berbasis pohon keputusan dan ensemble.

Selain itu, evaluasi performa model klasifikasi perlu dikaitkan secara langsung dengan metrik yang digunakan serta tujuan analisis yang ingin dicapai. *Pranckevičius* dan *Marcinkevičius* [21] menunjukkan bahwa pembahasan hasil klasifikasi yang komprehensif harus mencakup perbandingan antar model berdasarkan metrik evaluasi yang relevan, sehingga keunggulan dan keterbatasan masing-masing algoritma dapat dianalisis secara objektif. Oleh karena itu, pada bagian ini hasil penelitian disajikan secara bertahap, mulai dari deskripsi data, analisis spasial, evaluasi kinerja model, hingga pembahasan faktor-faktor utama yang memengaruhi tingkat strategis lokasi usaha *laundry* di Kota Surabaya.

### 3.1. Gambaran Umum Data

Penelitian ini menggunakan 801 titik data lokasi usaha *laundry* yang tersebar di 10 kecamatan di Kota Surabaya. Setiap titik mewakili satu usaha *laundry eksisting* dengan 22 variabel penentu yang terdiri dari faktor spasial, demografis, dan ekonomi. Data diperoleh melalui integrasi antara Google *Maps API*, OSM, dan BPS.

Karakteristik utama data ditunjukkan pada Tabel 1, yang memperlihatkan variabel-variabel yang digunakan dalam penelitian ini beserta jenis datanya.

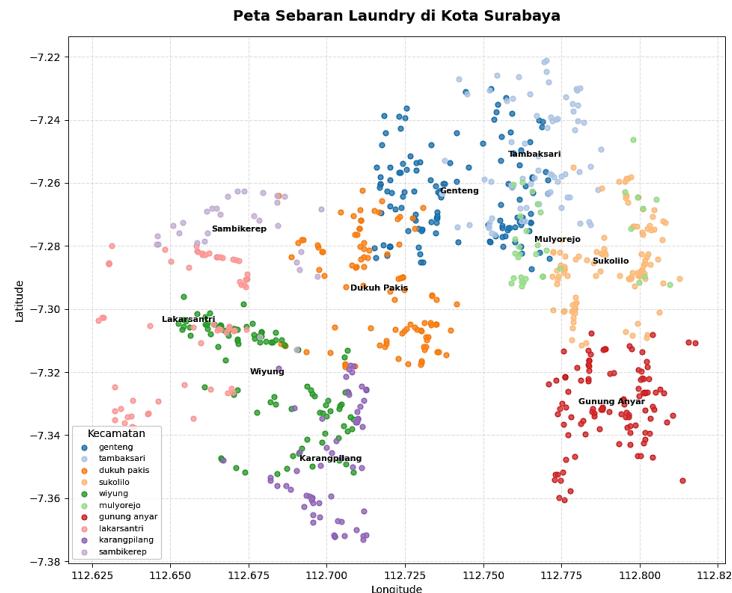
**Tabel 1.** Variabel Penelitian

Nama Variabel	Jenis Data	Sumber
Kepadatan Penduduk (jiwa/km <sup>2</sup> )	Numerik	BPS Kota Surabaya
Jumlah Kos di Radius 500 m	Numerik	Google Maps API
Jarak ke Permukiman (m)	Numerik	OSM
Jumlah Kompetitor (Laundry)	Numerik	Google Maps API
Harga Sewa Rata-rata (Rp/m <sup>2</sup> )	Numerik	OLX
Aksesibilitas Jalan	Kategorikal	OSM
Tingkat Ekonomi Wilayah	Kategorikal	BPS
Jumlah Penduduk	Numerik	BPS
Jarak ke Jalan Utama (m)	Numerik	OSM
Jarak ke Fasilitas Umum (m)	Numerik	Google Maps API

Sebagian besar lokasi usaha *laundry* terkonsentrasi di wilayah Surabaya Timur dan Pusat, seperti Kecamatan *Sukolilo*, *Mulyorejo*, dan *Genteng*, yang dikenal memiliki kepadatan penduduk tinggi serta tingkat mobilitas masyarakat yang aktif.

### 3.2. Analisis Deskriptif Spasial

Visualisasi awal data spasial ditampilkan pada Gambar 2, yang menunjukkan sebaran usaha *laundry eksisting* di Kota Surabaya. Dari peta tersebut terlihat bahwa wilayah dengan kepadatan penduduk tinggi dan jumlah rumah kos besar memiliki konsentrasi usaha *laundry* yang lebih padat.



**Gambar 2.** Persebaran lokasi usaha *laundry eksisting* di Surabaya

Pola persebaran ini mendukung teori *spatial clustering* di mana titik-titik usaha cenderung terkonsentrasi di area dengan potensi pasar tinggi [2]. Hal ini menunjukkan bahwa faktor spasial memainkan peran penting dalam menentukan peluang keberhasilan usaha *laundry* di wilayah perkotaan.

### 3.3. Hasil Pemodelan Klasifikasi

Untuk mengidentifikasi lokasi strategis dan tidak strategis, penelitian ini menerapkan lima algoritma klasifikasi: *Logistic Regression*, *Decision Tree*, *Naive Bayes*, *k-NN*, dan *Random Forest* [21]. Evaluasi kinerja dilakukan menggunakan *cross-validation* dengan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC*.

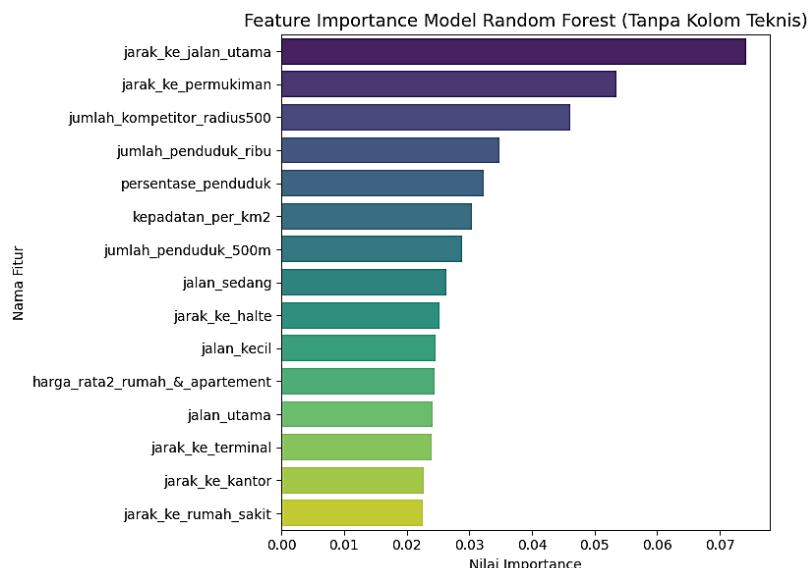
**Tabel 2.** Hasil perbandingan kinerja model klasifikasi

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
<i>Logistic Regression</i>	0.861	0.846	0.842	0.844	0.910
<i>Decision Tree</i>	0.874	0.859	0.867	0.863	0.922
<i>Naive Bayes</i>	0.811	0.795	0.782	0.788	0.864
k-NN	0.835	0.827	0.823	0.825	0.896
<i>Random Forest</i>	0.901	0.887	0.883	0.885	0.957

Berdasarkan Tabel 2, algoritma Random Forest menunjukkan kinerja terbaik dengan akurasi sebesar 90,1% dan nilai ROC-AUC 0,957. Model ini unggul karena mampu menangani data yang bersifat campuran (numerik dan *kategorikal*), serta memberikan feature importance yang dapat menjelaskan kontribusi tiap variabel terhadap hasil klasifikasi [1].

### 3.4. Analisis Faktor Penentu Lokasi Strategis

Hasil *feature importance* dari model *Random Forest* menunjukkan lima variabel yang paling berpengaruh dalam menentukan tingkat strategis lokasi *laundry* (lihat Gambar 3).

**Gambar 3.** Peringkat kepentingan variabel (*Feature Importance*)

Gambar 3 menampilkan peringkat kepentingan variabel (*feature importance*) yang dihasilkan oleh model *Random Forest*, yang menunjukkan kontribusi relatif setiap variabel dalam menentukan apakah suatu lokasi *laundry* tergolong strategis atau tidak strategis. Hasil tersebut menggambarkan bahwa kepadatan penduduk muncul sebagai variabel yang paling berpengaruh. Temuan ini konsisten dengan penelitian *geospasial* terdahulu yang menekankan bahwa konsentrasi penduduk berperan besar dalam menentukan potensi pasar dan intensitas permintaan jasa berbasis lokasi [3]. Wilayah dengan kepadatan penduduk yang tinggi cenderung memiliki aktivitas ekonomi yang lebih padat sehingga meningkatkan peluang layanan *laundry* untuk memperoleh pelanggan tetap.

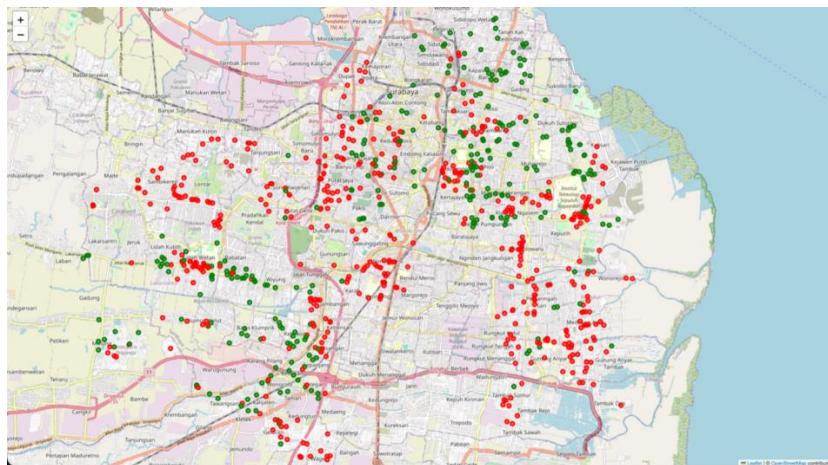
Variabel penting berikutnya adalah jumlah kompetitor, yang berperan sebagai indikator tingkat persaingan di suatu wilayah. *Random Forest* menunjukkan bahwa kompetisi memiliki hubungan non-linear terhadap kelayakan lokasi: lokasi dengan kompetitor dalam jumlah moderat menunjukkan nilai strategis yang lebih tinggi dibandingkan wilayah dengan kompetisi berlebih. Hal ini sejalan dengan temuan *Gao et al.* [4] mengenai perilaku spasial toko ritel, yang menyimpulkan bahwa pasar yang terlalu jenuh cenderung menurunkan performa bisnis baru. Sebaliknya, keberadaan beberapa kompetitor justru menandakan adanya permintaan yang stabil dan pasar yang aktif.

Selanjutnya, jarak ke permukiman juga menjadi faktor esensial. Lokasi *laundry* yang berdekatan dengan area pemukiman cenderung lebih mudah diakses oleh pelanggan, sehingga meningkatkan kemudahan mobilitas dan frekuensi kunjungan. Temuan ini konsisten dengan studi [10], yang menekankan bahwa aksesibilitas fisik terhadap permukiman merupakan salah satu variabel utama dalam pemodelan spasial untuk fasilitas layanan. Faktor berikutnya adalah harga sewa properti, yang memainkan peran penting dalam mempertimbangkan efisiensi biaya operasional. Biaya sewa yang terlalu tinggi umumnya mengurangi profitabilitas, namun harga sewa yang moderat pada wilayah dengan potensi permintaan tinggi dapat menjadi indikator lokasi yang menguntungkan [14].

Beberapa variabel lain seperti aksesibilitas jaringan jalan, jumlah rumah kos, serta jarak ke pusat aktivitas turut memiliki pengaruh yang signifikan meskipun tidak sebesar variabel utama. Temuan ini diperkuat oleh Ghosh & Das [16], yang menunjukkan bahwa konfigurasi ruang kota dan struktur permukiman memiliki dampak langsung terhadap konsentrasi aktivitas komersial. Secara keseluruhan, hasil *feature importance* ini menggambarkan pola spasial yang konsisten dengan literatur: kombinasi permintaan tinggi, kompetisi moderat, aksesibilitas baik, dan biaya sewa yang rasional menjadi faktor utama dalam menentukan lokasi usaha *laundry* yang strategis.

### 3.5. Visualisasi *Heatmap* dan Peta Klasifikasi

Untuk memperjelas hasil pemodelan, dibuat visualisasi spasial berupa *heatmap* lokasi strategis seperti ditunjukkan pada Gambar 4. Warna merah menandakan area strategis, sedangkan biru menunjukkan area yang kurang strategis.



**Gambar 4.** *Heatmap* distribusi lokasi strategis usaha *laundry*

Berdasarkan peta pada Gambar 4, wilayah Genteng, Sukolilo, dan Mulyorejo memiliki konsentrasi area strategis tertinggi. Wilayah ini didominasi oleh kepadatan penduduk tinggi, jumlah rumah kos banyak, serta kedekatan terhadap fasilitas umum seperti kampus, perumahan, dan perkantoran. Sementara itu, wilayah seperti Benowo dan Rungkut menunjukkan lebih banyak area non-strategis akibat jarak yang jauh dari pusat aktivitas dan kompetisi yang relatif rendah [6].

### 3.6. Pembahasan

Hasil penelitian menunjukkan bahwa pendekatan klasifikasi berbasis analisis *geospasial* mampu menilai kelayakan lokasi bisnis *laundry* dengan tingkat akurasi yang tinggi serta *interpretabilitas* yang baik. Model *Random Forest* menjadi model terbaik dibandingkan *Logistic Regression*, *Decision Tree*, *Naive Bayes*, dan k-NN. Keunggulan ini terutama disebabkan oleh kemampuannya dalam menangkap hubungan non-linier antara variabel spasial, demografis, dan ekonomi, serta dalam menghasilkan estimasi *feature importance* yang stabil pada data heterogen [7]. Variabel seperti kepadatan penduduk, jumlah kompetitor, jarak ke permukiman, dan harga sewa terbukti menjadi penentu kelayakan lokasi, dan pola kontribusi variabel ini sejalan dengan teori perilaku spasial urban yang menekankan bahwa interaksi antara permintaan dan intensitas kompetisi memengaruhi performa usaha jasa [3].

Jika dievaluasi lebih jauh, hasil klasifikasi menunjukkan bahwa lokasi laundry yang strategis cenderung berada pada kawasan dengan populasi padat, aksesibilitas tinggi, serta keberadaan kompetitor dalam jumlah moderat—suatu pola yang juga ditemukan pada penelitian Zhao et al. [3] mengenai lokasi coffee shop dan Gao et al. [4] dalam studi pemetaan distribusi toko ritel. Temuan ini mendukung konsep spatial clustering dan market threshold, yaitu bahwa usaha jasa seperti laundry membutuhkan cukup permintaan untuk bertahan, namun persaingan yang terlalu tinggi dapat menurunkan peluang keberhasilan. Selain itu, hubungan antara jarak permukiman dan performa usaha laundry menguatkan temuan Zheng et al. [2] yang menyatakan bahwa kedekatan fasilitas terhadap area hunian merupakan determinan utama dalam pemodelan kelayakan spasial.

Penerapan visualisasi *heatmap* dan peta interaktif juga memberikan wawasan tambahan mengenai distribusi spasial lokasi strategis. Pola konsentrasi area potensial yang terlihat pada visualisasi ini konsisten dengan penelitian Dong et al. [10] dan Yi & Huang [14] yang menekankan bahwa representasi spasial dapat mengungkap struktur permukiman dan aktivitas ekonomi yang tidak mudah terlihat dari analisis numerik semata. Dengan demikian, pendekatan Spatial Decision Support System (SDSS) yang dikembangkan dalam

penelitian ini telah berhasil memadukan model prediktif dan visualisasi spasial untuk memberikan pemahaman komprehensif bagi pelaku usaha maupun pengambil kebijakan.

Secara praktis, model klasifikasi yang dikembangkan dalam penelitian ini dapat membantu pelaku usaha *laundry* dalam menentukan lokasi ekspansi, menilai kelayakan lokasi potensial, dan meminimalkan risiko usaha berbasis posisi geografis. Pemerintah daerah dapat memanfaatkan hasil ini untuk mendukung perencanaan zonasi usaha mikro sehingga penyebaran layanan lebih merata dan efisien. Implikasi akademiknya adalah bahwa penelitian ini memperluas penerapan *machine learning* dalam domain pemodelan spasial UMKM, yang selama ini lebih banyak difokuskan pada usaha berskala besar seperti ritel modern atau restoran cepat saji [18].

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, model hanya mengandalkan variabel spasial, demografis, dan ekonomi yang tersedia dari data *open-source*, sehingga belum memasukkan faktor perilaku konsumen atau dinamika temporal seperti jam aktivitas harian. Hal ini serupa dengan keterbatasan yang ditemukan dalam penelitian urban *analytics* lainnya, seperti yang dijelaskan oleh Cheng & Adepeju [12] dan Xu *et al.* [8]. Kedua, penelitian ini hanya dilakukan pada satu kota, yaitu Surabaya, sehingga generalisasi ke kota lain memerlukan validasi ulang. Ketiga, meskipun *Random Forest* memberikan performa terbaik, model lain seperti *XGBoost* atau *Gradient Boosting* belum diuji, padahal penelitian-penelitian terbaru menunjukkan performa unggul pada data spasial non-linier [21].

Untuk pengembangan selanjutnya, penelitian dapat memperluas *dataset* dengan memasukkan variabel mobilitas penduduk, data transaksi usaha, maupun dinamika perilaku pelanggan untuk meningkatkan akurasi model. Selain itu, integrasi metode *deep learning* spasial seperti *Graph Neural Network* (GNN) dapat diuji sebagai alternatif dalam memahami hubungan antarwilayah secara lebih mendalam.

Secara keseluruhan, temuan penelitian ini tidak hanya memberikan kontribusi akademik dalam bidang analitik spasial dan *machine learning*, tetapi juga memiliki nilai terapan yang signifikan dalam mendukung proses pengambilan keputusan berbasis data di lingkungan perkotaan seperti Surabaya.

#### 4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi berbasis analisis *geospasial* untuk menilai tingkat strategis lokasi usaha *laundry* di Kota Surabaya dengan memanfaatkan data *open-source* yang berasal dari Google Maps API, OSM, OLX, dan BPS. Model dikembangkan melalui tahapan metodologis CRISP-DM yang meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Data Balancing*, *Modeling*, dan *Evaluation*. Hasil penelitian menunjukkan bahwa pendekatan klasifikasi dengan kombinasi data spasial, demografis, dan ekonomi mampu memberikan evaluasi yang objektif terhadap tingkat kelayakan lokasi usaha *laundry*. Dari lima model klasifikasi yang diuji, yaitu *Logistic Regression*, *Decision Tree*, *Naive Bayes*, k-NN, dan *Random Forest*, algoritma *Random Forest* menunjukkan performa terbaik dengan nilai akurasi 90,1% dan ROC-AUC 0,957.

Analisis *feature importance* pada model *Random Forest* mengungkapkan bahwa kepadatan penduduk, jumlah kompetitor dalam radius 500 meter, jarak ke permukiman, harga sewa rata-rata, dan jumlah rumah kos merupakan faktor-faktor paling berpengaruh terhadap klasifikasi lokasi strategis. Temuan ini menegaskan bahwa karakteristik spasial dan demografis wilayah memiliki peran yang sangat penting dalam menentukan peluang keberhasilan usaha *laundry* di kawasan perkotaan. Visualisasi hasil dalam bentuk *heatmap* dan peta klasifikasi memperlihatkan bahwa wilayah Genteng, Sukolilo, dan Mulyorejo merupakan area dengan potensi strategis tertinggi. Wilayah tersebut umumnya memiliki kepadatan penduduk tinggi, jumlah rumah kos banyak, serta kedekatan dengan fasilitas umum seperti kampus dan perkantoran. Sementara itu, wilayah Benowo dan Rungkut tergolong kurang strategis karena faktor jarak dan sebaran permukiman yang lebih jarang.

Secara keseluruhan, penelitian ini menunjukkan bahwa integrasi analisis spasial dan *machine learning* dapat menjadi pendekatan yang efektif untuk mendukung pengambilan keputusan bisnis berbasis data di sektor jasa. Model yang dihasilkan juga memiliki potensi untuk diterapkan secara lebih luas pada sektor lain yang bergantung pada lokasi, seperti ritel, logistik, dan kuliner, terutama di wilayah perkotaan yang padat penduduk.

#### REFERENCES

- [1] M. Georgati, H. S. Hansen, and C. Keßler, “Random Forest Variable Importance Measures for Spatial Dynamics: Case Studies from Urban Demography,” *ISPRS Int J Geoinf*, vol. 12, no. 11, p. 987, 2023, doi: 10.3390/ijgi12110987.
- [2] Z. Zheng, T. Morimoto, and Y. Murayama, “A GIS-Based Bivariate Logistic Regression Model for the Site-Suitability Analysis of Parcel-Pickup Lockers: A Case Study of Guangzhou, China,” *ISPRS Int J Geoinf*, vol. 10, no. 10, p. 871, 2021, doi: 10.3390/ijgi10100871.
- [3] J. Zhao, B. Zong, and L. Wu, “Site Selection Prediction for Coffee Shops Based on Multi-Source Space Data Using Machine Learning Techniques,” *ISPRS Int J Geoinf*, vol. 12, no. 8, p. 456, 2023, doi: 10.3390/ijgi12080456.

- [4] F. Gao *et al.*, “Location differs between traditional and new retail: A comparison analysis of Starbucks and Luckin Coffee using machine learning,” *Cities*, vol. 158, 2025, doi: 10.1016/j.cities.2024.104123.
- [5] M. D. Lagzi, S. M. Sajadi, and M. Taghizadeh-Yazdi, “A hybrid stochastic data envelopment analysis and decision tree for performance prediction in retail industry,” *Journal of Retailing and Consumer Services*, vol. 80, 2024, doi: 10.1016/j.jretconser.2024.103954.
- [6] J. Lu, X. Zheng, E. Nervino, Y. Li, Z. Xu, and Y. Xu, “Retail store location screening: A machine learning-based approach,” *Journal of Retailing and Consumer Services*, vol. 77, 2024, doi: 10.1016/j.jretconser.2023.103630.
- [7] T. Samal and A. Ghosh, “Ensemble-based predictive analytics for demand forecasting in multi-channel retailing,” *Expert Syst Appl*, vol. 299, p. 130212, Mar. 2026, doi: 10.1016/j.eswa.2025.130212.
- [8] Y. Chen, K. Xu, Y. He, and Q. Chen, “Utilitarian or experiential? Divergent location strategies of the rural coffee retail in the Hangzhou metropolitan area,” *Journal of Retailing and Consumer Services*, vol. 88, Jan. 2026, doi: 10.1016/j.jretconser.2025.104540.
- [9] J. Lu, X. Zheng, E. Nervino, Y. Li, Z. Xu, and Y. Xu, “Retail store location screening: A machine learning-based approach,” *Journal of Retailing and Consumer Services*, vol. 77, Mar. 2024, doi: 10.1016/j.jretconser.2023.103620.
- [10] L. Dong, J. Wu, and L. Zhang, “A hybrid machine learning approach for urban commercial zone classification,” *Comput Environ Urban Syst*, vol. 97, p. 101882, 2022, doi: 10.1016/j.compenvurbsys.2022.101882.
- [11] X. Zhu, Q. Li, and Y. Li, “Predicting optimal business locations from multi-source geospatial data,” *Land (Basel)*, vol. 12, no. 4, p. 845, 2023, doi: 10.3390/land12040845.
- [12] J. Cheng and M. Adepeju, “Spatial machine learning for point-of-interest demand estimation,” *Transactions in GIS*, vol. 26, no. 3, pp. 1159–1177, 2022, doi: 10.1111/tgis.12903.
- [13] J. Kang and H. Park, “POI-based business success prediction using ensemble classification,” *Expert Syst Appl*, vol. 218, p. 119556, 2023, doi: 10.1016/j.eswa.2022.119556.
- [14] C. Yi and B. Huang, “Spatial clustering and hotspot analysis for commercial site selection,” *Environment and Planning B*, vol. 48, no. 5, pp. 1243–1260, 2021, doi: 10.1177/2399808320930878.
- [15] Y. Li, J. Wang, and X. Huang, “Urban commercial hotspot prediction using ensemble geospatial models,” *Remote Sens (Basel)*, vol. 16, no. 3, p. 612, 2024, doi: 10.3390/rs16030612.
- [16] S. Ghosh and M. Das, “Assessment of service facility locations using GIS-based spatial analytics,” *GeoJournal*, vol. 85, pp. 1515–1534, 2020, doi: 10.1007/s10708-019-09992-y.
- [17] Y. Xu, S. Li, and X. Qiu, “Evaluating urban service accessibility using machine learning and GIS,” *Applied Geography*, vol. 122, p. 102246, 2020, doi: 10.1016/j.apgeog.2020.102246.
- [18] S. Hong, S. Lee, and J. Kim, “Retail location prediction using deep learning and geospatial big data,” *Sustainability*, vol. 13, no. 6, p. 3241, 2021, doi: 10.3390/su13063241.
- [19] D. Zhu, C. Cheng, W. Zhai, Y. Li, S. Li, and B. Chen, “Multiscale spatial polygonal object granularity factor matching method based on BPNN,” *ISPRS Int J Geoinf*, vol. 10, no. 2, Feb. 2021, doi: 10.3390/ijgi10020075.
- [20] D. Klimonová, V. Anderková, F. Babič, and L. T. Majnaric, “Decision Support System Improving the Interpretability of Generated Tree-Based Models,” *Acta Electrotechnica et Informatica*, vol. 22, no. 2, pp. 27–33, Jun. 2022, doi: 10.2478/aei-2022-0009.
- [21] T. Pranckevičius and V. Marcinkevičius, “Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification,” *Baltic Journal of Modern Computing*, vol. 5, no. 2, 2017, doi: 10.2236/bjmc.2017.5.2.05.