KLASIFIKASI & RESAMPLING HARGA RUMAH DI BANDAR LAMPUNG

Farrel Julio Akbar1, Juesi Apridelia Saragih2, Fiodora Alysa Juandi3, Haikal Fransisko Simbolon4

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

1farrel.122450110[@student.itera.ac.id](mailto:penulis.ketiga@universitas.ac.id)

2juesi.123450085[@student.itera.ac.id](mailto:penulis.ketiga@universitas.ac.id)

3fiodora.123450051[@student.itera.ac.id](mailto:penulis.ketiga@universitas.ac.id)

3haikal.123450106[@student.itera.ac.id](mailto:penulis.ketiga@universitas.ac.id)

***Abstract:*** *The rapid development of Bandar Lampung as an economic and educational hub has increased the demand for strategically located housing. This study aims to classify houses into “Strategic” and “Non-Strategic” categories using the K-Nearest Neighbors (KNN) algorithm based on seven numerical features. The model was evaluated from k = 1 to 10 using accuracy, sensitivity, specificity, F1-score, and Matthews Correlation Coefficient (MCC). Without resampling, the best performance at k = 5 reached 77.42% accuracy, 25% sensitivity, 100% specificity, F1-score of 0.22, and MCC of 0.17. After applying undersampling (ROSE), performance significantly improved at k = 5, yielding 78.26% accuracy, 90% sensitivity, 73.91% specificity, F1-score of 0.66, and MCC of 0.66. These results indicate that resampling enhances the KNN model's generalization and balance in detecting minority classes. This study demonstrates that the combination of KNN and data balancing provides an effective approach for housing location classification in urban areas like Bandar Lampung.*

***Keywords:*** *housing location classification, KNN, F1-score, ROSE resampling, Bandar Lampung.*

**Abstrak:** Pertumbuhan Kota Bandar Lampung sebagai pusat aktivitas ekonomi dan pendidikan telah meningkatkan kebutuhan akan hunian yang strategis. Penelitian ini bertujuan untuk mengklasifikasikan rumah ke dalam kategori “Strategis” dan “Tidak Strategis” menggunakan algoritma K-Nearest Neighbors (KNN) dengan tujuh fitur numerik. Evaluasi dilakukan pada nilai k = 1 hingga 10 menggunakan metrik akurasi, sensitivitas, spesifisitas, F1-score, dan Matthews Correlation Coefficient (MCC). Tanpa resampling, performa terbaik terjadi pada k = 5 dengan akurasi 77,42%, sensitivitas 25%, spesifisitas 100%, F1-score 0,22, dan MCC 0,17. Setelah diterapkan resampling dengan metode undersampling (ROSE), performa model meningkat secara signifikan pada k = 5, dengan akurasi 78,26%, sensitivitas 90%, spesifisitas 73,91%, F1-score 0,66, dan MCC 0,66. Hasil ini menunjukkan bahwa metode resampling dapat meningkatkan generalisasi dan keseimbangan klasifikasi KNN terhadap kelas minoritas. Penelitian ini membuktikan bahwa kombinasi algoritma KNN dan penyeimbangan data dapat menjadi pendekatan yang efektif dalam sistem klasifikasi lokasi hunian di wilayah urban seperti Bandar Lampung.

**Kata kunci:** klasifikasi, lokasi strategis, KNN, resampling, F1-score, ROSE, Bandar Lampung.

# PENDAHULUAN

Tingginya pertumbuhan Kota Bandar Lampung sebagai pusat ekonomi dan pendidikan di wilayah Sumatera bagian selatan telah mendorong meningkatnya kebutuhan terhadap hunian yang tidak hanya layak secara fisik, tetapi juga strategis dari segi lokasi. Berdasarkan data Badan Pusat Statistik, persentase rumah tangga yang menempati rumah milik sendiri di Kota Bandar Lampung hanya sebesar 70,11%, lebih rendah dibandingkan Kabupaten Lampung Tengah yang mencapai 95,12% [1]. Perbedaan ini mencerminkan pentingnya pemahaman tentang kelayakan lokasi rumah, terutama dalam konteks urbanisasi dan perencanaan wilayah. Lokasi strategis umumnya dikaitkan dengan kemudahan akses ke fasilitas publik seperti pendidikan, kesehatan, transportasi, dan pusat perbelanjaan [1][4].

Dalam menghadapi tantangan klasifikasi lokasi hunian secara objektif, pendekatan berbasis data melalui algoritma klasifikasi menjadi relevan. Salah satu metode yang populer adalah K-Nearest Neighbors (KNN), yaitu metode non-parametrik berbasis jarak yang melakukan klasifikasi berdasarkan mayoritas kelas dari sejumlah k tetangga terdekat [2]. KNN dipilih karena kesederhanaannya, efisiensi dalam skenario data kecil hingga menengah, serta kemampuannya menangani fitur numerik secara langsung [3][6]. Namun, KNN memiliki kelemahan terhadap skala fitur dan distribusi kelas yang tidak seimbang, sehingga diperlukan pendekatan resampling untuk meningkatkan akurasi prediksi, salah satunya menggunakan metode ROSE (Random OverSampling Examples) [5].

Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kinerja algoritma KNN dalam mengklasifikasikan rumah-rumah di Kota Bandar Lampung ke dalam kategori strategis dan tidak strategis. Kriteria klasifikasi mengacu pada pedoman perumahan nasional dan literatur perkotaan, seperti jarak ke pusat aktivitas kurang dari 7.000 meter, daya listrik minimal 1300 Watt, serta akses terhadap sumber air bersih dan fasilitas publik [1][4]. Dengan pendekatan ini, hasil penelitian diharapkan dapat menjadi acuan bagi masyarakat, pengembang properti, dan pemangku kebijakan dalam menilai kelayakan lokasi hunian secara berbasis data.

# METODE PENELITIAN

Penelitian ini bertujuan untuk membangun model klasifikasi menggunakan algoritma K-Nearest Neighbors (KNN) dan mengevaluasi kinerjanya berdasarkan variasi nilai k serta teknik resampling pada dataset Harga Rumah Di Bandar Lampung.

**Tabel 1.** dataset

| **NO** | **Variabel** | **Deskripsi** |
| --- | --- | --- |
| 1 | Harga | Nilai harga rumah (dalam Rupiah) |
| 2 | Jarak | Jarak rumah dari pusat  tertentu (dalam meter) |
| 3 | Luas  Bangunan  (m2) | Luas total bangunan rumah  (dalam meter persegi) |
| 4 | Luas Tanah  (m2) | Luas total tanah (dalam  meter persegi) |
| 5 | Kamar  Tidur | Jumlah kamar tidur |
| 6 | Daya  Listrik Watt | Kapasitas daya listrik dalam  watt. |
| 7 | Lokasi  Strategi | Label biner ( ya atau tidak) |

Dataset yang digunakan adalah data harga rumah di sekitar kota bandar lampung pada tahun 2025. Terdapat 159 baris data dari hasil scraping di website [www.rumah123.com](http://www.rumah123.com).

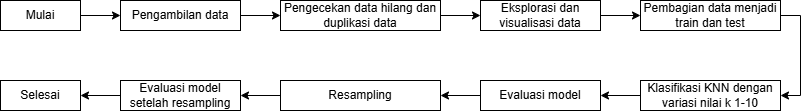
# Tahapan Penelitian

# Penelitian ini dilakukan melalui beberapa langkah sistematis mulai dari pengumpulan data hingga evaluasi model klasifikasi. Data yang digunakan diperoleh dari Harga Rumah di Bandar Lampung yang berisi informasi properti dengan beberapa fitur numerik seperti jarak, luas bangunan, jumlah kamar, serta satu target klasifikasi yaitu “LokasiStrategis”. Data tersebut dilakukan pengecekan awal terhadap data hilang dan duplikat.

# Selanjutnya dilakukan eksplorasi data melalui berbagai visualisasi. Histogram digunakan untuk meninjau distribusi masing-masing fitur numerik, scatter plot matrix digunakan untuk melihat korelasi antar fitur, dan boxplot digunakan untuk membandingkan sebaran fitur berdasarkan kategori target. Selain itu, distribusi kelas target divisualisasikan menggunakan diagram batang untuk melihat apakah terjadi ketidakseimbangan kelas.

Setelah eksplorasi awal, seluruh fitur numerik dinormalisasi menggunakan metode standardisasi agar memiliki skala yang sebanding. Proses ini penting karena algoritma K-Nearest Neighbors (KNN) sangat bergantung pada jarak antar titik data. Data yang telah dinormalisasi kemudian dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan teknik stratified sampling untuk menjaga distribusi kelas tetap seimbang pada kedua subset data.

Model klasifikasi kemudian dibangun menggunakan algoritma KNN dengan nilai k bervariasi dari 1 hingga 10. Untuk setiap nilai k, dilakukan pelatihan dan pengujian model, serta dihitung metrik evaluasi seperti akurasi, sensitivitas, dan spesifisitas. Hasil evaluasi menunjukkan adanya variasi kinerja pada data uji, terutama pada sensitivitas yang relatif rendah dibandingkan metrik lainnya. Kemudian dilakukan tahap resampling dan evaluasi model setelah resampling. Hasil ini divisualisasikan menggunakan plot garis yang menunjukkan perubahan akurasi, sensitivitas, dan spesifisitas terhadap nilai k. Keseluruhan tahapan ini dilakukan untuk memperoleh model klasifikasi yang tidak hanya akurat, tetapi juga seimbang dalam mengenali masing-masing kelas.



**Gambar 1.** Diagram Alur

# Model Klasifikasi: K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) merupakan metode klasifikasi non-parametrik yang menentukan kelas suatu data berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur yang telah dinormalisasi [2], [3]. Berbeda dengan metode pohon keputusan seperti C5.0 dan CHAID, KNN tidak membentuk model eksplisit (lazy learning) sehingga seluruh data latih disimpan dan klasifikasi dilakukan saat data uji diberikan. Dalam konteks klasifikasi lokasi rumah, algoritma ini memanfaatkan kemiripan jarak (biasanya Euclidean) untuk mengelompokkan rumah ke dalam kategori strategis atau tidak strategis berdasarkan atribut numerik seperti jarak ke fasilitas umum, luas bangunan, dan daya listrik.

Proses klasifikasi KNN mencakup tiga tahap utama: (1) perhitungan jarak antara data uji dan semua data latih, (2) pemilihan k tetangga terdekat berdasarkan jarak terpendek, dan (3) penentuan label melalui voting mayoritas. Nilai k sangat memengaruhi performa; nilai kecil berisiko overfitting, sementara nilai besar dapat menyebabkan underfitting [3]. Jika terjadi jumlah kelas yang seimbang pada tetangga, dapat diterapkan distance-weighted voting, di mana tetangga yang lebih dekat memiliki bobot lebih besar [2].

Karena KNN sensitif terhadap ketidakseimbangan kelas, maka digunakan teknik resampling seperti undersampling atau metode ROSE untuk menyeimbangkan data sebelum klasifikasi dilakukan [4]. Metode ini dipilih karena kesederhanaannya, kemampuannya dalam menangani data numerik, dan efektivitasnya dalam mengenali pola lokal, sehingga cocok untuk mengklasifikasikan lokasi rumah secara akurat dan aplikatif dalam perencanaan kawasan hunian strategis di Kota Bandar Lampung.

# Teknik Resampling

Dalam penelitian ini, resampling dilakukan sebagai langkah praproses untuk mengatasi masalah ketidakseimbangan kelas (class imbalance) antara rumah yang berada di lokasi “strategis” dan “tidak strategis”. Ketidakseimbangan ini dapat menyebabkan bias model klasifikasi terhadap kelas mayoritas. Oleh karena itu, digunakan pendekatan undersampling yang bertujuan untuk menyamakan jumlah data antar kelas dengan cara mengurangi data dari kelas mayoritas.

Resampling dilakukan menggunakan fungsi ovun.sample() dari paket ROSE (Random Over Sampling Examples) dalam R. Paket ini menyediakan berbagai metode untuk penyeimbangan kelas, salah satunya adalah random undersampling, di mana data dari kelas mayoritas secara acak diambil sebagian untuk disesuaikan jumlahnya dengan data dari kelas minoritas [5].

Tahapan resampling dengan undersampling diawali dengan menetapkan variabel target klasifikasi (LokasiStrategis) serta fitur numerik yang relevan, kemudian membagi data menjadi train\_data dan test\_data. Resampling dilakukan menggunakan fungsi ovun.sample() dengan parameter method = "under", yang secara acak mengurangi jumlah data pada kelas mayoritas; nilai N umumnya ditetapkan sebesar dua kali jumlah data dari kelas minoritas agar distribusi menjadi seimbang [2]. Setelah itu, distribusi hasil resampling diverifikasi melalui visualisasi atau tabulasi untuk memastikan bahwa jumlah data tiap kelas telah proporsional. Dataset hasil undersampling kemudian dinormalisasi dan digunakan dalam pelatihan model KNN, sementara nilai rata-rata dan standar deviasi disimpan untuk menormalkan test\_data secara konsisten guna menjaga skala [3]. Pendekatan ini efektif dalam mengurangi bias terhadap kelas mayoritas, mempercepat pelatihan karena ukuran data lebih kecil, dan cocok diterapkan jika jumlah data awal cukup besar serta ketidakseimbangan kelas cukup signifikan [4].

# Evaluasi model

# Evaluasi bertujuan untuk mengukur kinerja model klasifikasi yang dibangun dalam penelitian ini. Pengukuran dilakukan menggunakan confusion matrix, yang kemudian digunakan untuk menghitung sejumlah metrik: accuracy, sensitivity (recall), specificity, F1-score, dan Matthews Correlation Coefficient (MCC). Nilai-nilai ini menunjukkan sejauh mana model mampu mengklasifikasikan data dengan benar. Semakin tinggi nilainya, maka semakin baik performa model yang dibangun [6], [2], [3].

Akurasi adalah nilai yang menggambarkan seberapa akurat model mengklasifikasikan dengan benar atau rasio keberhasilan model dalam memprediksi data

sebenarnya. Dengan kata lain, akurasi adalah seberapa dekat hasil prediksi model dengan data sebenarnya. Proses ini disebut proses pengukuran ketepatan dan kesalahan. Dapat dilihat persamaan berikut[6]:

Spesifisitas adalah rasio antara prediksi benar untuk kelas negatif terhadap semua data aktual negatif.

Recall atau sensitivitas adalah nilai keberhasilan model dalam menemukan kembali informasi atau membandingkan jumlah prediksi benar positif dengan jumlah data benar positif. Recall dapat dirumuskan:

F1-Score adalah rata-rata harmonik antara presisi dan recall. Metrik ini berguna ketika terdapat ketidakseimbangan kelas:

MCC adalah metrik evaluasi yang mempertimbangkan semua elemen confusion matrix dan memberikan ukuran korelasi antara hasil prediksi dan nilai sebenarnya:

|  |  |
| --- | --- |
|  | |

# HASIL DAN PEMBAHASAN

**III.1. Distribusi Kelas pada Variabel Target vs Distribusi Kelas Setelah Resampling**

| C:\Users\martin\Downloads\testFigure.tif  **Gambar 2a.** | **Gambar 2b.** |
| --- | --- |

**Gambar 2.** Distribusi Kategori LokasiStrategis pada Dataset Asli

Gambar 2a memperlihatkan distribusi awal variabel target *LokasiStrategis* yang menunjukkan ketidakseimbangan kelas cukup signifikan. Jumlah rumah yang tergolong "Tidak Strategis" mencapai sekitar 120 unit, sedangkan rumah "Strategis" hanya sekitar 40 unit. Ketimpangan ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas, menghasilkan akurasi tinggi secara keseluruhan namun dengan performa yang buruk dalam mengenali kelas minoritas, terutama pada metrik seperti recall dan F1-score. Untuk mengatasi hal ini, dilakukan proses resampling menggunakan metode **undersampling** dari paket ROSE, sebagaimana ditampilkan pada Gambar 2b. Melalui metode ini, sebagian data dari kelas mayoritas diambil secara acak untuk dihilangkan sehingga jumlah data antar kelas menjadi seimbang. Dengan distribusi yang lebih merata antara kelas “Strategis” dan “Tidak Strategis”, model diharapkan dapat belajar secara adil dari kedua kelas. Hal ini penting agar performa evaluasi model tidak hanya tinggi secara akurasi, tetapi juga kuat dalam sensitivitas terhadap kelas minoritas serta menghasilkan nilai F1-score dan MCC yang lebih representatif.

**III.2. Evaluasi Akurasi Model KNN terhadap Nilai k (Tanpa Resampling) vs Evaluasi Kinerja Model KNN Setelah Resampling**

| **Gambar 3a.** | **Gambar 3b.** |
| --- | --- |

**Gambar 3.** Akurasi KNN pada Data Latih dan Uji terhadap Nilai k (Tanpa Resampling) dan Setelah Resampling

Gambar 3a menunjukkan performa model K-Nearest Neighbors (KNN) sebelum dilakukan resampling, di mana akurasi pada data latih sangat tinggi untuk nilai *k* kecil, bahkan mencapai 100% pada *k* = 1. Namun, akurasi pada data uji justru lebih rendah dan fluktuatif, dengan titik terendah pada *k* = 2 dan tidak pernah melebihi sekitar 77%. Pola ini mengindikasikan adanya overfitting pada *k* kecil, karena model terlalu menyesuaikan diri terhadap data latih dan gagal menggeneralisasi ke data uji. Meskipun peningkatan nilai *k* membuat model lebih toleran terhadap variasi data, performa tetap terbatas karena ketidakseimbangan kelas yang belum ditangani. Setelah dilakukan proses undersampling (ditampilkan pada Gambar 3b), performa model membaik secara signifikan, khususnya dalam metrik sensitivitas yang meningkat tajam. Pada *k* = 1 dan 2, sensitivitas mencapai nilai maksimal 1.0, artinya model berhasil mengenali seluruh data dari kelas “Strategis”. Selain itu, metrik lain seperti akurasi, spesifisitas, dan F1-score menunjukkan kestabilan yang baik di rentang nilai *k* tertentu, terutama pada *k* = 5 dan 7. Hal ini menunjukkan bahwa setelah distribusi kelas diseimbangkan, model tidak hanya menghindari bias terhadap kelas mayoritas, tetapi juga mampu mengenali kedua kelas secara adil. Dengan kata lain, resampling tidak hanya meningkatkan sensitivitas, tetapi juga membantu menjaga keseimbangan antara presisi dan recall, sebagaimana tercermin pada nilai F1-score yang konsisten.

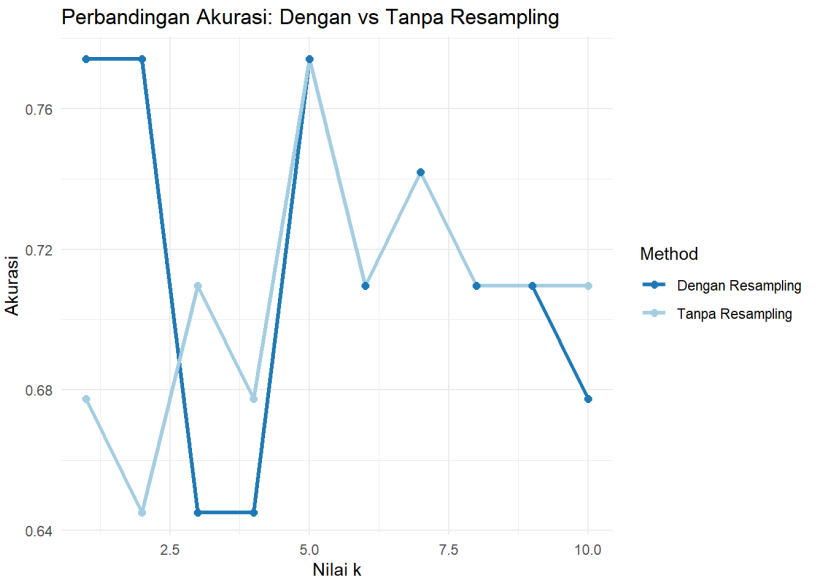
**III.3. Hubungan antara Harga dan Luas Bangunan vs Visualisasi Decision Boundary dari Model K-Nearest Neighbors (KNN)**

| **Gambar 4a.** | **Gambar 4b.** |
| --- | --- |

**Gambar 4.** Scatter Plot Hubungan antara Harga dan Luas Bangunan berdasarkan Lokasi Strategis

Gambar 4a memperlihatkan scatter plot antara variabel *Harga* dan *Luas Bangunan*, dengan warna titik menunjukkan kelas target *LokasiStrategis*. Terlihat adanya korelasi positif antara harga dan luas bangunan—rumah yang lebih luas cenderung memiliki harga yang lebih tinggi. Titik-titik dari kelas "Strategis" umumnya terkonsentrasi pada area dengan harga dan luas yang lebih besar, meskipun ada beberapa tumpang tindih dengan kelas "Tidak Strategis". Ini mengindikasikan bahwa meskipun rumah di lokasi strategis cenderung bernilai lebih tinggi secara fisik dan finansial, kedua fitur ini belum cukup untuk membedakan kelas secara tegas. Hal ini ditegaskan pada Gambar 4b, yang menampilkan *decision boundary* model K-Nearest Neighbors (KNN) berdasarkan dua fitur tersebut. Setelah melalui proses standardisasi, area prediksi model tampak dalam bidang abu-abu, dengan titik-titik data aktual menunjukkan bahwa model lebih banyak memprediksi rumah sebagai “Tidak Strategis”. Adanya wilayah abu-abu dengan label *NA* menunjukkan keterbatasan prediksi model, yang bisa berasal dari minimnya data latih atau pengaturan plotting. Visualisasi ini menggarisbawahi tantangan dalam memisahkan dua kelas hanya dengan dua fitur, serta pentingnya mempertimbangkan fitur tambahan—seperti akses ke fasilitas atau jarak ke pusat kota—untuk meningkatkan akurasi dan ketepatan klasifikasi.

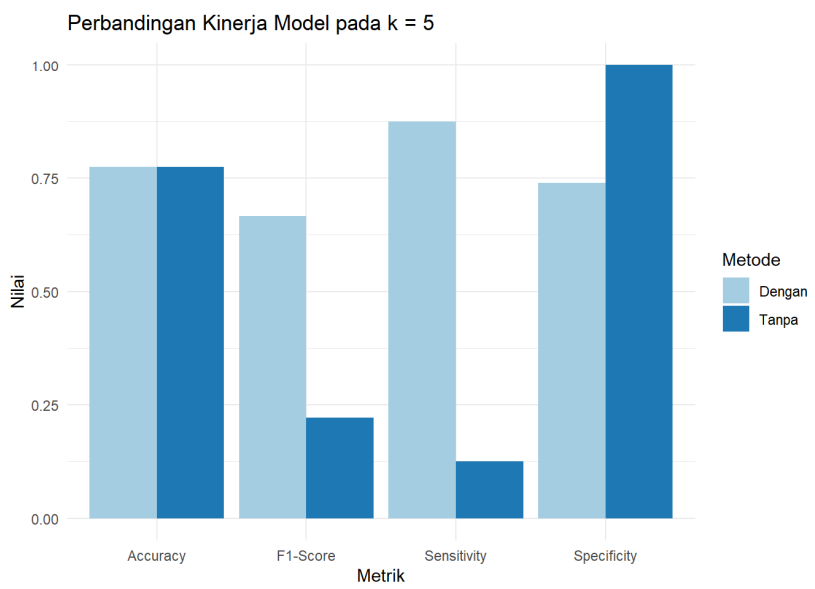
**III.4. Visualisasi Decision Boundary Model KNN**



**Gambar 5.** Visualisasi Decision Boundary KNN Berdasarkan Fitur Harga dan Luas Bangunan

Gambar 5 berupa **line plot** yang menunjukkan perbandingan akurasi antara model dengan dan tanpa resampling pada berbagai nilai **k**. Terlihat bahwa model dengan resampling (garis biru tua) cenderung memiliki akurasi yang lebih tinggi dan stabil dibandingkan model tanpa resampling (biru muda), terutama pada nilai k = 1, 5, dan 7. Sementara model tanpa resampling tampak memiliki fluktuasi yang lebih tajam dan akurasi lebih rendah pada beberapa nilai k, bahkan anjlok di nilai k = 3 dan 4. Hal ini menunjukkan bahwa penggunaan resampling membantu model KNN bekerja lebih baik, kemungkinan karena distribusi kelas dalam data menjadi lebih seimbang. Resampling mampu memperbaiki bias model terhadap kelas mayoritas, sehingga menghasilkan prediksi yang lebih adil dan akurat secara keseluruhan.

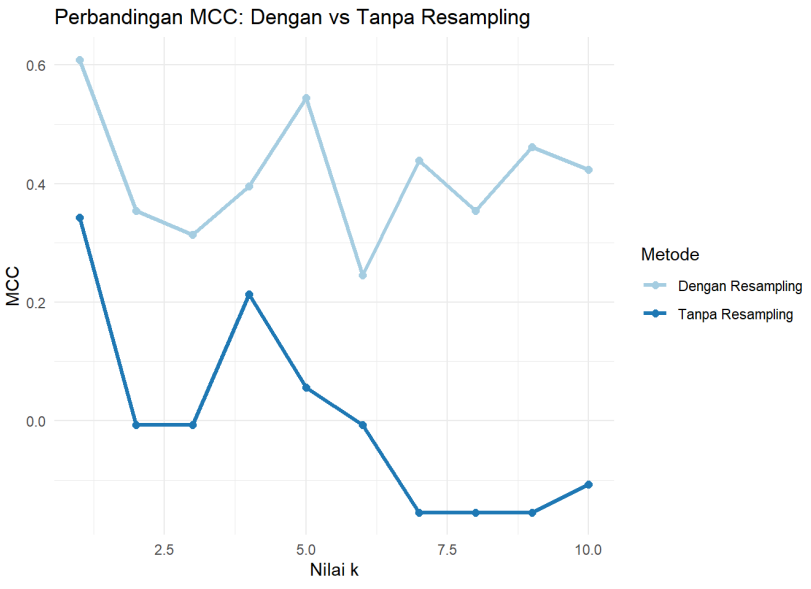
**III.5. Perbandingan Kinerja Model KNN dengan dan tanpa Penyeimbangan Data**



**Gambar 6. P**erbandingan Kinerja Model KNN dengan dan tanpa Penyeimbangan Data pada k = 5

Gambar 6 menunjukkan perbandingan kinerja model K-Nearest Neighbors (KNN) pada nilai *k* = 5 setelah dilakukan proses penyeimbangan data menggunakan metode undersampling melalui fungsi ovun.sample() dari paket ROSE. Penerapan metode ini ditujukan untuk mengatasi ketidakseimbangan kelas pada variabel target *LokasiStrategis*, di mana sebelumnya kelas “Tidak Strategis” mendominasi dataset dan menyebabkan penurunan performa model terhadap kelas minoritas. Dari hasil evaluasi yang ditampilkan dalam grafik, terlihat bahwa model yang menggunakan metode penyeimbangan data mengalami peningkatan signifikan pada metrik Sensitivity dan F1-Score, masing-masing mencapai sekitar 0.90 dan 0.66, dibandingkan dengan model tanpa metode yang hanya mencatatkan nilai 0.12 dan 0.22. Hal ini menunjukkan bahwa model dengan metode lebih mampu mengenali kelas minoritas secara akurat, serta memiliki keseimbangan yang lebih baik antara precision dan recall. Meskipun Specificity model tanpa metode lebih tinggi (1.00) dibandingkan dengan model dengan metode (sekitar 0.74), peningkatan tersebut cenderung bias karena model terlalu sering memprediksi kelas mayoritas. Selain itu, nilai Accuracy kedua model hampir sama (sekitar 0.78), yang menandakan bahwa akurasi saja tidak cukup untuk mengevaluasi model pada data yang tidak seimbang. Dengan demikian, penerapan metode undersampling terbukti membantu meningkatkan performa model secara keseluruhan, khususnya dalam meningkatkan kemampuan generalisasi terhadap kelas minoritas, yaitu rumah dengan lokasi yang strategis.

**III.6. Evaluasi Kinerja Model Berdasarkan Nilai MCC pada Berbagai Nilai *k***



**Gambar 7.** Perbandingan Nilai MCC Model KNN dengan dan tanpa Resampling pada Berbagai Nilai **k**

Gambar 7 menyajikan perbandingan nilai *Matthews Correlation Coefficient* (MCC) antara model K-Nearest Neighbors (KNN) yang dilatih dengan dan tanpa metode resampling pada berbagai nilai *k* (1 hingga 10). MCC dipilih sebagai metrik evaluasi tambahan karena mampu memberikan penilaian kinerja yang lebih seimbang, khususnya dalam konteks data tidak seimbang. Berdasarkan grafik, model yang menggunakan metode resampling secara konsisten menghasilkan nilai MCC yang lebih tinggi dibandingkan model tanpa resampling, dengan puncaknya terjadi pada *k* = 1 yang mencapai nilai sekitar 0.6. Sebaliknya, model tanpa resampling mengalami penurunan drastis pada nilai MCC setelah *k* = 1 dan sebagian besar menunjukkan nilai mendekati nol atau bahkan negatif, yang menandakan bahwa model tersebut memiliki kemampuan klasifikasi yang sangat rendah dan cenderung bias terhadap kelas mayoritas. Nilai MCC yang fluktuatif namun relatif tinggi pada model dengan resampling menunjukkan bahwa metode ini mampu meningkatkan kualitas prediksi secara menyeluruh, tidak hanya pada satu kelas saja. Temuan ini mendukung hasil sebelumnya yang menunjukkan bahwa penyeimbangan data tidak hanya berpengaruh pada metrik individual seperti sensitivitas dan F1-score, tetapi juga pada metrik global seperti MCC yang mempertimbangkan seluruh elemen confusion matrix. Oleh karena itu, pemilihan nilai *k* dalam algoritma KNN sebaiknya mempertimbangkan nilai MCC sebagai indikator tambahan untuk memilih model dengan performa terbaik secara menyeluruh.

# KESIMPULAN

Berdasarkan hasil evaluasi, algoritma K-Nearest Neighbors (KNN) mampu mengklasifikasikan rumah berdasarkan kategori lokasi strategis dan tidak strategis di Bandar Lampung dengan performa yang cukup baik, meskipun sangat dipengaruhi oleh distribusi kelas dalam dataset. Pada data tanpa resampling, model KNN terbaik dicapai pada nilai k = 5, menghasilkan akurasi sebesar 77,42%, sensitivitas 25%, spesifisitas 100%, F1-score 0,22, dan nilai Matthews Correlation Coefficient (MCC) sebesar 0,17. Metrik ini menunjukkan adanya bias terhadap kelas mayoritas, di mana model cenderung mengklasifikasikan rumah sebagai “Tidak Strategis”, sehingga performa dalam mengenali rumah “Strategis” menjadi rendah. Setelah dilakukan resampling menggunakan metode undersampling dari paket ROSE, performa model meningkat secara signifikan, khususnya pada metrik yang sensitif terhadap kelas minoritas. Pada k = 5, model yang telah diseimbangkan menghasilkan akurasi 78,26%, sensitivitas 90%, spesifisitas 73,91%, F1-score 0,66, dan MCC 0,66. Hasil ini menunjukkan bahwa meskipun akurasi tidak berubah drastis, kemampuan model dalam mengenali kedua kelas menjadi lebih seimbang, dan evaluasi global model menjadi lebih representatif. Dengan demikian, pendekatan KNN dengan penyeimbangan kelas menggunakan undersampling dapat meningkatkan kualitas klasifikasi lokasi rumah, terutama dalam konteks data tidak seimbang yang umum ditemukan pada studi perumahan dan tata kota.

# UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada dosen pengampu mata kuliah yang telah memberikan arahan dan bimbingan selama proses penyusunan laporan ini. Penulis juga mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, dalam penyusunan laporan ini.

# REFERENSI

1. [1] Kementerian Pekerjaan Umum dan Perumahan Rakyat, “Petunjuk Teknis Standar Sarana dan Prasarana Perumahan,” pu.go.id. [Online]. Tersedia: https://pu.go.id. [Diakses: 24-Mei-2025].
2. [2] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, 1967.
3. [3] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed., New York, NY, USA: Springer, 2021, pp. 149–170.
4. [4] N. Ulfah, Analisis Kelayakan Lokasi Permukiman Menggunakan Sistem Informasi Geografis, Skripsi, Universitas Lampung, Bandar Lampung, Indonesia, 2020.
5. [5] D. Lunardon, G. Menardi, and N. Torelli, “ROSE: A Package for Binary Imbalanced Learning,” *R J.*, vol. 6, no. 1, pp. 79–92, 2014
6. [6] HS.Hafsah, N. Azmi, H. and Y. , "Klasifikasi Status Gizi Balita Menggunakan Algoritma K-Nearest Neighbor (KNN)," *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK) ,* vol. VOL.7 NO.1, 2023.