

# BAB I

## PENDAHULUAN

Bab pendahuluan pada laporan penelitian ini memberikan penjelasan mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang digunakan pada dokumen ini yang berjudul *Prediksi Harga Rumah Menggunakan Metode DBSCAN dan Random Forest pada Data Iklan Properti*.

### 1.1 Latar Belakang

Properti merupakan salah satu sektor penting dalam ekonomi negara. Sektor ini memiliki kontribusi yang besar dalam pertumbuhan ekonomi. Selain itu, sektor properti juga berdampak pada ketersediaan serta keterjangkauan rumah atau tempat tinggal bagi masyarakat. Di sisi lain, harga properti seperti rumah juga menjadi salah satu topik yang umum diperbincangkan oleh masyarakat (Hamami & Dahlan, 2024; Wang dkk., 2019).

Harga rumah bergantung ke banyak faktor, di antaranya adalah karakteristik fisik bangunan, lokasi, kondisi lingkungan sekitar, serta persepsi masyarakat terhadap faktor-faktor tersebut. Di samping itu, harga suatu aset juga dipengaruhi oleh tingkat kesediaan masyarakat untuk membayarnya. Banyaknya faktor yang memengaruhi tersebut membuat penentuan atau prediksi terhadap harga rumah menjadi tugas sulit dan memakan banyak waktu. Kompleksnya prediksi harga rumah membuat otomatisasi terhadap proses tersebut dapat menjadi salah satu jawaban yang efektif. Dengan begitu, otomatisasi terhadap prediksi harga rumah dapat memberikan manfaat bagi pihak-pihak di bidang properti, seperti pembeli, penjual, notaris, serta pembuat kebijakan. Salah satu manfaat tersebut adalah memberikan wawasan mengenai kondisi pasar properti, seperti rumah dengan lebih baik (Geerts dkk., 2023; Hernes dkk., 2024).

Otomatisasi prediksi harga rumah dapat dilakukan dengan algoritma *machine learning*. Pada prediksi harga sebuah aset, terdapat dua jenis strategi yang dapat diterapkan. Strategi pertama adalah melakukan prediksi terhadap tren harga pasar dalam format deret waktu, contohnya memprediksi harga saham selama beberapa periode tertentu. Strategi kedua adalah melakukan prediksi terhadap suatu aset berdasarkan karakteristik yang dimilikinya, contohnya memprediksi harga rumah. Pada penelitian ini, strategi yang

diterapkan adalah strategi kedua, yaitu prediksi harga aset berdasarkan karakteristiknya (Adetunji dkk., 2021).

Penelitian terdahulu menunjukkan bahwa *machine learning* sudah banyak digunakan untuk melakukan prediksi terhadap harga rumah atau properti residensial lain. Hamami dan Dahlan (2024) melakukan perbandingan terhadap beberapa metode dalam memprediksi harga rumah di Pulau Jawa. Hasilnya menunjukkan bahwa algoritma *Random Forest Regression* menunjukkan kestabilan yang tinggi dan memiliki nilai *Mean Squared Error* (MSE) paling kecil, yaitu 0,000681. Hernes dkk. (2024) juga mengembangkan model prediksi harga *real estate* residensial di Polandia. Pada penelitian tersebut, hasilnya menunjukkan bahwa meskipun *Gradient Boosting Regression* mampu mendapatkan nilai *R-Squared* ( $R^2$ ) tertinggi, yaitu 0,989, tetapi *Random Forest Regression* mampu mendapatkan nilai  $R^2$  yang sangat kompetitif, yaitu 0,986. Penelitian serupa juga dilakukan oleh Quang dkk. (2020), yaitu implementasi model *regression* untuk memprediksi harga rumah pada *dataset* “Housing Price in Beijing”. Hasilnya menunjukkan bahwa *Stacked Generalization Regression* mampu mendapatkan nilai *Root Mean Squared Logarithmic Error* (RLMSE) terendah, yaitu 0,16350. Meskipun demikian, *Random Forest Regression* juga memiliki performa yang tinggi dan nilai RLMSE yang didapatkan adalah 0,16568, hanya berbeda 1,3333%.

Beberapa penelitian menunjukkan bahwa model *Random Forest Regression* cukup umum digunakan dalam kasus prediksi harga rumah. Selain itu, model tersebut juga konsisten dalam mendapatkan akurasi tinggi. Dengan begitu, model *Random Forest Regression* dapat dijadikan model dasar untuk melakukan prediksi terhadap harga rumah. Oleh karena itu, penelitian ini melakukan prediksi harga rumah berdasarkan karakteristiknya dengan menggunakan model *Random Forest Regression*.

Penelitian mengenai prediksi harga rumah menggunakan *machine learning* yang dilakukan oleh Hamami dan Dahlan (2024) dan Quang dkk. (2020) juga melibatkan metode *preprocessing* pada data. Metode tersebut diterapkan karena menurut Bala & Behal (2024), *preprocessing* pada data dapat memberikan pengaruh performa yang signifikan pada model *machine learning*. Maka, penelitian ini juga menerapkan metode *preprocessing* sebelum model *machine learning* dilatih. Salah satu metode *preprocessing* yang penting untuk dilakukan adalah penghapusan terhadap data yang terdeteksi sebagai *outliers* (Zou, 2022).

*Outliers* merupakan nilai ekstrem yang memiliki perbedaan besar dengan nilai-nilai lain pada sekumpulan nilai, sehingga keberadaannya jarang (Rajeswari dkk., 2018). Salah satu metode yang diusulkan oleh Majeed dan Hamad (2023) untuk mendeteksi *outliers* pada data yang memiliki dimensi tinggi adalah menerapkan PCA untuk menurunkan dimensi data, kemudian menggunakan metode *clustering*, yaitu DBSCAN untuk memberikan label pada *outliers*. DBSCAN telah digunakan oleh beberapa penelitian untuk mendeteksi *outliers*, contohnya penelitian yang dilakukan oleh Gnanadurai dan Vinothina (2023) menerapkan DBSCAN untuk mendeteksi *outliers* pada data yang diperoleh dari *wireless sensor networks*. Lalu, Monalisa (2018) juga menerapkan DBSCAN pada penelitiannya untuk mendeteksi *outliers* guna menganalisis loyalitas pelanggan. Dengan demikian, metode yang digunakan pada penelitian ini untuk *outliers cleaning* adalah DBSCAN.

Selain metode *preprocessing*, *hyperparameter tuning* juga merupakan proses yang dapat memengaruhi performa model *machine learning*. Hal ini karena model *machine learning*, seperti model *Random Forest* memiliki sejumlah parameter statis, yang disebut dengan *hyperparameter*. Parameter-parameter ini merupakan konfigurasi dari model *machine learning* yang telah ditetapkan sebelum proses pelatihan dan tidak akan berubah nilainya selama pelatihan (Dabool dkk., 2024). Oleh karena itu, pada penelitian ini, proses *hyperparameter tuning* turut diimplementasikan

## **1.2 Rumusan Masalah**

Adapun rumusan masalah dalam penelitian ini berdasarkan latar belakang yang telah disusun adalah bagaimana performa kombinasi metode DBSCAN sebagai metode *outliers cleaning* dan model *Random Forest* dalam memprediksi harga rumah pada data iklan properti.

## **1.3 Tujuan dan Manfaat**

Tujuan dari penelitian ini adalah mengetahui performa model prediksi dengan mengombinasikan DBSCAN sebagai metode *outliers cleaning* dan model *Random Forest* dalam melakukan prediksi harga rumah berdasarkan data iklan properti.

Manfaat dari penelitian ini adalah tersedianya model prediksi yang dapat memprediksi harga rumah berdasarkan data iklan properti. Model yang tersedia tersebut diharapkan dapat digunakan sebagai referensi untuk pengembangan model prediksi harga rumah yang lebih akurat.

#### **1.4 Ruang Lingkup**

Pemerincian ruang lingkup dimaksudkan agar penelitian yang dilakukan lebih terarah dan mencapai tujuan yang diharapkan. Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Data iklan properti yang digunakan pada penelitian ini adalah hasil dari penarikan data dari situs web penyedia iklan properti, yaitu Lamudi. Pada situs tersebut dilakukan penyaringan iklan properti dengan hanya menyertakan properti tipe rumah yang dijual di Kabupaten pada Provinsi Jawa Barat. Setelah dilakukan penyaringan, data iklan rumah akan diambil, termasuk karakteristik, harga, dan lokasinya. Data iklan tersebut dikumpulkan dan diformat menjadi CSV.
2. Waktu penarikan data dari iklan rumah tersebut adalah tanggal 10 Maret 2025.
3. Data iklan rumah yang berhasil didapatkan adalah sejumlah 41.656 baris.
4. Data iklan rumah hasil penarikan tersebut belum melalui proses validasi untuk memastikan bahwa harga rumah yang tercantum di situs web adalah harga penawaran yang sesungguhnya.

#### **1.5 Sistematika Penulisan**

Struktur penulisan dalam laporan penelitian ini terdiri dari lima bab yang membahas penelitian “Prediksi Harga Rumah Menggunakan Metode DBSCAN dan *Random Forest* pada Data Iklan Properti”. Laporan penelitian ini ditulis dalam lima bab agar penulisan menjadi sistematis serta mudah dipahami. Berikut ini adalah ringkasan pembahasan dari setiap bab:

## BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan laporan penelitian.

## BAB II TINJAUAN PUSTAKA

Bab ini menyajikan studi pustaka tentang landasan teori yang dijadikan acuan dalam penelitian ini: *state-of-the-art*, Pengertian Rumah, *Machine Learning (ML)*, *Decision Tree*, *Random Forest*, *Data Preprocessing*, *K-Fold Cross Validation*, evaluasi model, serta *tools* dan *library* yang digunakan.

## BAB III METODE PENELITIAN

Bab ini menguraikan metodologi yang digunakan dalam penelitian ini, yaitu garis besar penelitian, proses penarikan dan pengubahan format data, melakukan *preprocessing* pada data, pembangunan model prediksi menggunakan model *Random Forest*, dan evaluasi pada model yang telah dibangun. Pada bagian akhir, dilakukan perancangan aplikasi prediksi harga rumah menggunakan data iklan rumah yang telah melalui tahap *preprocessing*.

## BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan pembahasan dan analisis dari hasil penelitian. Hal yang disajikan tersebut meliputi pembahasan hasil performa model *Random Forest* dalam memprediksi harga rumah. Model tersebut dilatih menggunakan *K-Fold Cross Validation*. Pada bagian akhir, terdapat pembahasan mengenai implementasi aplikasi prediksi yang telah dirancang.

## BAB V PENUTUP

Bab ini berisi kesimpulan dari semua bab yang telah dibahas sebelumnya serta saran untuk pengembangan penelitian selanjutnya.