

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jumlah penduduk di kota Jakarta selalu mengalami peningkatan dalam beberapa tahun terakhir (BPS, 2023). Pada tahun 2021, tercatat sebanyak 10.605.437 jiwa penduduk yang tinggal di Kota Jakarta. Angka ini kemudian meningkat menjadi 10.640.007 jiwa di tahun 2022, menunjukkan kenaikan sebesar 0,33%. Tren peningkatan ini berlanjut di tahun 2023, di mana jumlah penduduk Kota Jakarta mencapai 10.672.100 jiwa, atau naik 0,30% dibandingkan tahun 2022. Pertumbuhan penduduk Jakarta yang pesat disebabkan oleh implementasi kota ini sebagai pusat pemerintahan dan perekonomian di Indonesia. Kondisi ini menarik para pendatang dari berbagai daerah di Indonesia untuk mencari pekerjaan dan penghidupan yang lebih layak, hal tersebut mempengaruhi peningkatan jumlah penduduk di Jakarta.

Jumlah penduduk yang terus meningkat membawa dampak besar pada kebutuhan dasar manusia. Salah satu kebutuhan primer yang paling terpengaruh adalah papan, atau tempat tinggal. Rumah adalah salah satu kebutuhan dari masyarakat yang tidak dapat dihindari karena rumah merupakan kebutuhan primer, tempat untuk berlindung, tempat untuk beristirahat dari penatnya aktifitas harian (Saiful et al., 2021). Terutama di kota-kota besar seperti Jakarta, Jakarta merupakan kota metropolitan dan memiliki penduduk yang beragam yang terdiri dari berbagai suku, agama, dan latar belakang budaya. Kondisi demografi dan perekonomian masyarakat yang heterogen memberikan pengaruh terhadap harga jual rumah di Jakarta. Sehingga harga rumah yang ditawarkan semakin bervariasi.

Ada beberapa faktor yang mempengaruhi harga rumah. Dalam penelitiannya Rahadi et al., (2015) membagi faktor-faktor tersebut menjadi tiga kelompok utama, yaitu kondisi fisik, konsep, dan lokasi. Kondisi fisik merupakan sifat-sifat yang dimiliki suatu rumah yang dapat diamati dengan indera manusia, antara lain ukuran rumah, jumlah kamar tidur, jumlah kamar mandi, ketersediaan dapur dan garasi, ketersediaan taman, luas tanah dan bangunan, serta usia rumah (Limsombunchai, 2004), sedangkan konsep merupakan ide yang ditawarkan oleh pengembang yang dapat menarik calon pembeli, misalnya konsep rumah minimalis, lingkungan sehat

dan hijau, serta lingkungan elit. Lokasi merupakan faktor penting dalam membentuk harga sebuah rumah. Hal ini karena lokasi menentukan harga tanah yang berlaku (Deng & Kang, 2013).

Di era digital ini, kemajuan teknologi telah memberikan akses yang lebih mudah kepada calon pembeli rumah untuk menelusuri berbagai penawaran rumah di situs *online*. Beragamnya harga rumah yang tercantum di situs tersebut menjadi pertimbangan tersendiri bagi para calon pembeli. Berbagai faktor seperti spesifikasi, lokasi, dan fasilitas di sekitar rumah dapat sangat memengaruhi harga sebuah rumah. Salah satu platform daring yang bisa dimanfaatkan untuk mencari rumah adalah OLX.co.id. OLX.co.id merupakan situs online yang fokus pada layanan barang dan jasa, termasuk iklan properti seperti rumah. Keberadaan situs ini sangat membantu banyak orang dalam mendapatkan informasi tentang rumah yang dijual, sehingga memudahkan mereka dalam memilih rumah yang sesuai dengan kebutuhan.

Prediksi harga rumah merupakan salah satu upaya penting bagi kedua pihak, calon pembeli dan penjual. Bagi pembeli rumah, Sharma et al., (2024). Bagi penjual rumah, akan sangat bermanfaat untuk memahami perkiraan harga rumah di pasar untuk mengajukan harga penawaran yang tepat (H. Zhang et al., 2023). Metode yang dapat digunakan untuk melakukan identifikasi/prediksi tersebut adalah dengan melakukan analisis regresi.

Analisis regresi adalah metode yang memanfaatkan hubungan antara dua variabel atau lebih. Analisis regresi terbagi menjadi dua berdasarkan banyaknya jumlah variabel bebas yang digunakan, Regresi Linier Sederhana dimana hanya melibatkan satu variabel bebas terhadap variabel terikat, dan Regresi Linier Berganda yang melibatkan dua atau lebih variabel bebas terhadap variabel terikat. Analisis regresi dilakukan untuk mengetahui korelasi antara variabel bebas dan variabel terikat yang mempunyai hubungan sebab-akibat, dan membuat prediksi dengan menggunakan hubungan tersebut (Uyanik & Güler, 2013). Selain dapat dilakukan dengan metode klasik, analisis regresi dapat juga dilakukan dengan pendekatan *Machine learning* (ML).

Machine learning (ML) adalah metode analisis data yang berguna untuk melakukan suatu tugas klasifikasi maupun prediksi. ML merupakan bidang ilmu

yang mempelajari algoritma komputer di mana algoritma tersebut secara otomatis dapat semakin baik kinerjanya seiring dengan pengalaman dan data yang dipakai. ML merupakan metode berbasis data yang menggabungkan konsep dasar dalam ilmu komputer dengan ide-ide dari statistik, probabilitas, dan optimasi (Mohri et al., 2018). Untuk melakukan klasifikasi atau prediksi, algoritma ML membangun model berdasarkan data sampel, yang juga disebut sebagai data pelatihan lalu menerapkan algoritma tersebut pada data pengujian. ML dapat dibagi ke beberapa tipe utama, tergantung pada cara belajarnya. Pada algoritma *Unsupervised learning*, data masukan yang digunakan tidak memiliki label. *Unsupervised learning* belajar untuk menemukan pola dan struktur tersembunyi pada keseluruhan data tersebut. Sedangkan pada *Supervised learning*, data yang diberikan sudah memiliki label. *Supervised learning* memungkinkan untuk menghasilkan *output* data yang didapat dari pengalaman sebelumnya (Crisci et al., 2012). Analisis regresi dengan pendekatan ML dianggap menghasilkan prediksi yang lebih akurat dibandingkan regresi dengan pendekatan klasik berdasarkan data jumlah data yang digunakan (Fitri, 2020).

Salah satu algoritma ML yang dapat digunakan untuk memprediksi harga rumah adalah *Random Forest* (RF). RF adalah termasuk suatu *ensemble learning* dimana teknik dalam pembelajaran mesin yang menggabungkan prediksi beberapa model untuk menciptakan model prediksi yang lebih baik. *Ensemble learning* pada RF adalah dengan menggunakan teknik *bootstrap aggregating* (bagging) adalah teknik yang digunakan dalam *ensemble learning*, bagging mengambil banyak sampel dengan penggantian (Jatmiko et al., 2019). Dengan melatih subkumpulan data yang berbeda, algoritma cenderung tidak mempelajari pola yang sama dari *noise* (pencilan) tertentu dalam data, sehingga dapat mengurangi varians model.

RF adalah algoritma ML yang dapat digunakan untuk tugas klasifikasi jika variabel bertipe kategorik dan tugas regresi jika variabel kontinu. RF merupakan algoritma yang menggunakan konsep dasar pohon keputusan (*decision tree/DT*). RF bekerja dengan menggunakan banyak DT untuk melakukan analisis, dimana hasil akhirnya berdasarkan *majority vote* (suara terbanyak) saat tugas klasifikasi dan rata-rata prediksi semua pohon saat tugas regresi (Brownlee, 2016). Pada RF tugas regresi adalah masalah memprediksi nilai sebenarnya untuk setiap item, atau dapat

disebut juga *random forest regression* (RFR). RFR memiliki keunggulan utama, yaitu dapat mengatasi *overfitting* dengan *feature randomness*/fitur acak, yaitu saat menciptakan banyak pohon dari data yang berbeda, RFR hanya mempertimbangkan subset fitur acak. Hal ini mencegah pohon untuk terlalu fokus pada fitur spesifik yang mungkin terjadi secara kebetulan dalam data pelatihan namun tidak dapat digeneralisasikan (Breiman, 2001). Walaupun begitu, RF memiliki kekurangan. Salah satunya adalah cenderung menghasilkan model yang sulit untuk diinterpretasi. Ini karena RF menggunakan banyak pohon yang digabungkan untuk membuat keputusan akhir, sehingga semakin sulit untuk menentukan pohon mana yang paling banyak berkontribusi pada membuat keputusan akhir.

Berdasarkan penelitian sebelumnya, *Random Forest Regression* (RFR) menunjukkan performa yang baik dalam melakukan prediksi. (Eren, 2023) membandingkan beberapa metode dalam melakukan analisis prediksi, menunjukan nilai akurasi prediksi menggunakan metode RFR mendapat skor evaluasi *R-squared* (R^2) tertinggi yaitu 0.92 sedangkan *Multilayer Perceptron* $R^2 = 0.65$, *K-nearest neighbor regression* $R^2 = 0.47$, dan *Support vector regression* $R^2 = 0.20$. Haryanto et., al (2023) menunjukkan metode RFR memiliki performa baik dengan nilai akurasi 81,6% dibandingkan dengan Regresi Linier Berganda dengan nilai akurasi sebesar 78,5%. Tanamal., et al (2023) menunjukan metode *Random Forest* mendapat skor prediksi yang tinggi yaitu 88% dibandingkan *Decision Tree* 75%, *K-Nearest Neighbour* 70%, *Logistic* dan *Support Vector Machine* masing-masing 65%.

Dengan demikian penulis tertarik melakukan prediksi harga rumah dengan metode *Random forest regression*, dimana masih menjadi suatu hal yang menarik untuk diteliti karena berbagai variabel yang terlibat semakin beragam, sehingga harga rumah berubah dari waktu ke waktu, khususnya pada kota besar seperti Jakarta.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka penelitian ini memiliki beberapa permasalahan yang dapat dirumuskan sebagai berikut:

1. Bagaimana penerapan RFR dalam memprediksi harga rumah di Jakarta dibandingkan dengan Regresi Linier Berganda?

2. Faktor apa saja yang dianggap penting dalam memprediksi harga rumah di Jakarta berdasarkan metode *random forest regression*?

1.3 Batasan Masalah

Berdasarkan rumusan masalah di atas, maka batasan masalah pada penelitian ini metode *ensemble* yang digunakan hanya *random forest regression* dengan teknik *bagging*. Dan optimasi *hyperparameter* yang digunakan, yaitu *GridSearchCV*.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, maka tujuan dilakukannya penelitian ini adalah:

1. Menerapkan metode RFR dan Regresi Linier Berganda untuk memprediksi harga rumah di Jakarta.
2. Membandingkan kinerja RFR dengan metode Regresi Linier Berganda dalam memprediksi harga rumah di Jakarta.
3. Mengidentifikasi tingkat kepentingan faktor-faktor yang mempengaruhi harga rumah di Jakarta.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat digunakan sebagai metode alternatif untuk memprediksi, serta mengetahui faktor apa saja yang dianggap penting dalam memprediksi harga rumah di wilayah Jakarta, sehingga membantu calon pembeli dalam menentukan anggaran dan memilih rumah yang sesuai dengan kemampuan finansial mereka, serta bagi penjual dapat menentukan harga jual yang kompetitif dan memaksimalkan keuntungan dari penjualan rumah mereka.