

Implementasi Regresi Linear Berganda dengan Metode Normal Equation untuk Prediksi Harga Properti

Muhamad Dimas Saputra

Teknik Informatika/Sains dan Teknologi

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

Tangerang Selatan, Indonesia

muhamad.dimas24@mhs.uinjkt.ac.id

Abstract—Sektor properti memiliki peran vital dalam ekonomi, namun penentuan harga rumah seringkali menjadi tantangan kompleks karena dipengaruhi oleh banyak faktor fisik. Penelitian ini bertujuan untuk mengimplementasikan metode Regresi Linear Berganda menggunakan algoritma *Gradient Descent* yang dibangun secara manual (*from scratch*) untuk memprediksi harga rumah. Data yang digunakan berasal dari dataset perumahan standar yang telah melalui proses pra-pemrosesan dan standarisasi. Hasil pengujian menunjukkan bahwa model mampu meminimalkan *Mean Squared Error* (MSE) secara iteratif menuju konvergensi.

Index Terms—Prediksi Harga Rumah, Regresi Linear, Gradient Descent, Machine Learning, Metode Numerik.

I. PENDAHULUAN

A. Latar Belakang

Sektor properti memegang peranan vital dalam perekonomian global, namun prediksi harga properti tetap menjadi tantangan yang kompleks karena dipengaruhi oleh berbagai faktor seperti ukuran bangunan, jumlah kamar, lokasi, dan kondisi properti itu sendiri. Rumah tidak hanya berfungsi sebagai kebutuhan pokok untuk tempat berlindung, tetapi juga dianggap sebagai bentuk investasi masa depan yang strategis karena adanya fluktuasi harga yang terus meningkat seiring waktu. Oleh karena itu, diperlukan metode yang efektif untuk memprediksi harga properti secara akurat guna mendukung pengambilan keputusan bagi pembeli, penjual, maupun investor.

Perkembangan teknologi telah mendorong penggunaan metode berbasis *machine learning* untuk menyelesaikan masalah prediksi ini, di mana komputer dapat belajar dari data historis tanpa harus diprogram secara eksplisit. Salah satu pendekatan yang terbukti efektif adalah Regresi Linear Berganda (*Multiple Linear Regression*). Penelitian sebelumnya menunjukkan bahwa metode ini memungkinkan pemahaman yang lebih mendalam mengenai keterkaitan antara variabel fisik bangunan—seperti luas tanah dan luas bangunan—with nilai jual rumah, serta mampu memberikan prediksi dengan tingkat akurasi yang memuaskan [1].

Secara matematis, tujuan dari metode ini adalah memilih parameter bobot (w) yang dapat meminimalkan fungsi biaya

(*Cost Function*) atau selisih kuadrat antara prediksi dan nilai aktual. Untuk menyelesaikan masalah optimasi ini, digunakan algoritma *Gradient Descent*, yang bekerja secara iteratif memperbarui parameter dengan mengambil langkah ke arah penurunan tercuram (*steepest decrease*) dari fungsi biaya hingga mencapai konvergensi [2].

Selain pemilihan algoritma, tahapan pra-pemrosesan data juga sangat krusial. Data mentah seringkali memiliki skala yang berbeda-beda yang dapat menghambat kinerja model. Penelitian Hallan dan Fajri menekankan bahwa proses standarisasi fitur (*feature standardization*) sangat penting untuk memastikan konsistensi data, yang pada akhirnya membantu algoritma berbasis *gradient descent* bekerja lebih efisien dan menghasilkan prediksi yang lebih akurat [3].

Berdasarkan landasan tersebut, penelitian ini akan mengimplementasikan Regresi Linear Berganda dengan optimasi *Gradient Descent* yang dibangun dari awal (*from scratch*) untuk memprediksi harga properti.

B. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

- 1) Bagaimana memodelkan hubungan antara variabel independen (luas bangunan, jumlah kamar, dan jarak lokasi) terhadap harga rumah menggunakan persamaan fungsi Regresi Linear:

$$Y = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n \quad (1)$$

- 2) Bagaimana penerapan algoritma *Gradient Descent* secara iteratif untuk memperbarui parameter bobot w guna meminimalkan fungsi biaya ($J(w)$)?
- 3) Seberapa besar pengaruh tahapan pra-pemrosesan data, khususnya standarisasi fitur, dalam meningkatkan efisiensi dan akurasi model regresi linear?

C. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- 1) Membangun model prediksi harga properti menggunakan algoritma *Linear Regression* untuk menganalisis pengaruh berbagai faktor fisik terhadap harga.

- 2) Mengimplementasikan algoritma optimasi *Gradient Descent* secara manual untuk mencari nilai minimum dari fungsi *Least Squares*.
- 3) Mengevaluasi kinerja model yang dihasilkan menggunakan metrik *Mean Squared Error* (MSE) untuk mengukur rata-rata kesalahan prediksi antara nilai aktual dan nilai prediksi.

D. Batasan Masalah

Agar pembahasan lebih terarah, penulis menetapkan batasan masalah sebagai berikut:

- **Metode Algoritma:** Penelitian ini berfokus pada penggunaan algoritma Regresi Linear Berganda yang diselesaikan menggunakan metode numerik *Gradient Descent*, bukan menggunakan solusi analitik tertutup (*Normal Equations*).
- **Variabel Data:** Variabel yang digunakan untuk prediksi dibatasi pada fitur fisik dan lokasi yang relevan, seperti luas bangunan, jumlah kamar, dan area terkait, sebagaimana divalidasi signifikansinya dalam penelitian terdahulu.
- **Teknik Pra-pemrosesan:** Data akan melalui proses standarisasi (*StandardScaler*) sebelum pelatihan untuk menangani perbedaan skala antar fitur, guna mempercepat konvergensi algoritma.
- **Implementasi:** Kode program dibangun menggunakan bahasa Python dengan operasi matriks dasar (NumPy), tanpa menggunakan fungsi pelatihan instan dari *library* pembelajaran mesin tingkat tinggi.

II. LANDASAN TEORI

A. Regresi Linear Berganda

Regresi Linear Berganda (*Multiple Linear Regression*) adalah metode statistik yang digunakan untuk memodelkan hubungan linear antara satu variabel dependen (target) dengan dua atau lebih variabel independen (fitur). Dalam konteks penelitian ini, variabel dependen adalah harga rumah, sedangkan variabel independen adalah karakteristik fisik rumah seperti luas tanah, jumlah kamar, dan lokasi [1].

Secara matematis, jika terdapat n fitur input x_1, x_2, \dots, x_n , maka fungsi prediksi Y didefinisikan sebagai kombinasi linear dari bobot parameter w :

$$Y(x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (2)$$

Dimana:

- $Y(x)$ adalah nilai prediksi (harga rumah).
- w_0 adalah bias (*intercept*).
- w_1, \dots, w_n adalah koefisien bobot (*weights*) untuk setiap fitur.

Untuk menyederhanakan notasi, kita dapat mendefinisikan $x_0 = 1$, sehingga persamaan dapat ditulis dalam bentuk vektor:

$$Y(x) = w^T x \quad (3)$$

Dimana w dan x adalah vektor berdimensi \mathbb{R}^{n+1} .

B. Fungsi Biaya (Cost Function)

Untuk mengevaluasi seberapa baik model memprediksi data, diperlukan fungsi objektif yang disebut Fungsi Biaya (*Cost Function*). Penelitian ini menggunakan *Mean Squared Error* (MSE) yang mengukur rata-rata kuadrat selisih antara nilai prediksi (Y) dan nilai aktual (y).

Berdasarkan teori *Least Squares* yang dijelaskan oleh Ng dan Ma, fungsi biaya $J(w)$ didefinisikan sebagai [2]:

$$J(w) = \frac{1}{2m} \sum_{i=1}^m (Y^{(i)} - y^{(i)})^2 \quad (4)$$

Dimana:

- m adalah jumlah data pelatihan (*training examples*).
- $Y^{(i)}$ adalah hasil prediksi untuk data ke- i .
- $y^{(i)}$ adalah label harga asli untuk data ke- i .
- Konstanta $\frac{1}{2}$ digunakan untuk memudahkan perhitungan turunan nantinya.

Tujuan dari pelatihan model adalah mencari nilai vektor w yang meminimalkan $J(w)$.

C. Algoritma Gradient Descent

Karena tidak menggunakan solusi analitik tertutup (*Normal Equations*), pencarian nilai minimum $J(w)$ dilakukan menggunakan metode numerik iteratif bernama *Gradient Descent* (Algoritma LMS). Algoritma ini bekerja dengan cara memperbarui nilai parameter w secara bertahap ke arah yang berlawanan dengan gradien fungsi biaya [2].

Aturan pembaruan (*update rule*) untuk setiap parameter w_j adalah:

$$w_j := w_j - \alpha \frac{\partial}{\partial w_j} J(w) \quad (5)$$

Dimana α adalah *learning rate* yang mengontrol ukuran langkah setiap iterasi. Turunan parsial dari fungsi biaya terhadap bobot w_j dapat diturunkan menjadi:

$$\frac{\partial}{\partial w_j} J(w) = \frac{1}{m} \sum_{i=1}^m (Y^{(i)} - y^{(i)}) x_j^{(i)} \quad (6)$$

Proses ini diulang (*looping*) hingga nilai fungsi biaya konvergen atau mencapai batas iterasi yang ditentukan.

D. Pra-pemrosesan Data (Standardisasi)

Sebelum data dilatih menggunakan *Gradient Descent*, sangat penting untuk melakukan penskalaan fitur (*Feature Scaling*). Hal ini dikarenakan variabel input memiliki rentang nilai yang sangat berbeda (misalnya, luas rumah dalam ratusan meter persegi, sedangkan jumlah kamar hanya satuan digit).

Penelitian Hallan dan Fajri menunjukkan bahwa tanpa standarisasi, kontur fungsi biaya akan berbentuk ellips yang sangat lonjong, yang menyebabkan algoritma *Gradient Descent* lambat mencapai konvergensi atau berosilasi [3]. Teknik yang digunakan adalah *Z-score Normalization*:

$$x'_j = \frac{x_j - \mu_j}{\sigma_j} \quad (7)$$

Dimana μ_j adalah rata-rata (mean) dari fitur j , dan σ_j adalah standar deviasi dari fitur j . Dengan teknik ini, setiap fitur akan memiliki rata-rata 0 dan standar deviasi 1, sehingga proses optimasi menjadi lebih efisien.

III. METODOLOGI

A. Pra-pemrosesan Data

Sebelum dilakukan pelatihan, data mentah harus melalui tahap pra-pemrosesan. Salah satu tahapan krusial adalah standarisasi fitur (*Feature Scaling*) menggunakan teknik *Z-score normalization*:

$$x' = \frac{x - \mu}{\sigma} \quad (8)$$

Hal ini penting agar kontur fungsi biaya menjadi lebih simetris, sehingga *Gradient Descent* dapat konvergen lebih cepat [3].

IV. HASIL DAN PEMBAHASAN

Implementasi dilakukan menggunakan bahasa Python dengan pustaka NumPy. Grafik penurunan *Loss* dapat dilihat pada Gambar 1.

Fig. 1. Ilustrasi Grafik Penurunan Cost Function (Contoh Placeholder)

V. KESIMPULAN

Berdasarkan hasil eksperimen, algoritma Regresi Linear dengan optimasi Gradient Descent terbukti mampu memprediksi harga rumah dengan tingkat error yang dapat diterima.

REFERENCES

- [1] R. N. T. Siregar, V. Sitorus, and W. P. Ananta, "Analisis prediksi harga rumah di bandung menggunakan regresi linear berganda," *Journal of Creative Student Research (JCSR)*, vol. 1, no. 6, pp. 395–404, 2023.
- [2] A. Ng and T. Ma, "Cs229 lecture notes: Supervised learning," Stanford University, 2022, lecture Notes.
- [3] R. R. Hallan and I. N. Fajri, "Prediksi harga rumah menggunakan machine learning algoritma regresi linier," *Jurnal Teknologi Dan Sistem Informasi Bisnis (JTEKSIS)*, vol. 7, no. 1, pp. 57–62, 2025.