

Penggunaan Least Square Regression untuk Memprediksi Kejatuhan Harga Mobil Bekas

Maxwell Zefanya Ginting
2306221200

Departemen Teknik Elektro
Universitas Indonesia
Depok, Indonesia
maxwell.zefanya@ui.ac.id

Abstract—Mobil adalah salah satu jenis komoditas yang dianggap sebagai barang mewah. Lantas, harga dari mobil pun jatuh relatif lebih cepat bila dibandingkan dengan harga jenis kendaraan lainnya, seperti motor. Tujuan utama ditulisnya studi kasus ini mencoba untuk memprediksi bagaimana harga mobil bekas jatuh menggunakan least square regression.

Keywords—Komputasi Numerik, Least Square, Regresi, Regresi Polinomial

I. PENDAHULUAN

Mobil adalah salah satu jenis komoditas yang dianggap sebagai barang mewah. Lantas, harga dari mobil pun jatuh relatif lebih cepat pada 2-5 tahun pertama bila dibandingkan dengan harga jenis kendaraan lainnya, seperti motor. Dampaknya, banyak orang yang ingin membeli mobil pertama kali lebih memilih untuk membeli mobil bekas ketimbang mobil yang masih baru. Selain harga, salah satu faktor yang menjadi penentu pembeli untuk membeli sebuah model mobil adalah ketahanan harga dari mobil tersebut. Bila harga mobil jatuh pesat beberapa tahun setelah pembelian, misal 5 tahun, maka pembeli akan mencoba untuk melihat model mobil yang lainnya. Studi kasus ini akan mencoba untuk menggunakan regresi Least Square dalam aplikasi prediksi kejatuhan harga mobil bekas.

Secara definisi, least square regression adalah metode regresi yang mencoba untuk menemukan fungsi yang paling sesuai terhadap kumpulan titik data yang diberikan. Fungsi ini bisa berbentuk banyak hal, seperti linear, polynomial, dan lain-lain. Metode ini bekerja dengan data yang tidak sepenuhnya akurat. Artinya berbeda dengan interpolasi, metode regresi tidak akan menghasilkan fungsi yang melewati semua titik data dengan tepat. Hal ini relevan dengan studi kasus karena data yang diberikan adalah harga mobil bekas. Harga jual dari mobil bekas ditentukan oleh banyak faktor, dan bukan hanya tahun produksinya saja. Supaya bisa menghasilkan fungsi prediksi yang berguna, maka fakta tersebut harus diterima. Itulah sebabnya kenapa studi kasus ini akan menggunakan metode regresi.

II. STUDI LITERATUR

Pada bagian ini, akan dijelaskan beberapa definisi yang akan digunakan berikutnya pada studi kasus kali ini.

A. Harga Jual

Harga jual didefinisikan sebagai berapa harga jual mobil bekas yang ditawarkan oleh penjual. Harga jual memiliki satuan Rupiah (Rp.)

B. Tahun Produksi

Tahun Produksi didefinisikan sebagai tahun berapa mobil tersebut diproduksi sesuai dengan yang tertera pada STNK. Tahun produksi sering dijadikan sebagai patokan awal berapa harga jual sebuah mobil.

C. Least Square Regression

Least Square Regression (selanjutnya disebut sebagai regresi) adalah metode statistik dasar yang digunakan untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen. Regresi bisa dibedakan menjadi banyak jenis, tergantung hasil fungsi akhir yang ingin didapatkan. Pada studi kasus ini, regresi yang dipakai adalah regresi polynomial. Lebih tepatnya, regresi polynomial orde tiga. Artinya, hasil fungsi akhir yang dikeluarkan oleh metode ini akan ada dalam bentuk sebagai berikut.

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + e \quad (1)$$

Di mana y adalah variabel dependen data, x adalah variabel independen pada data, a_0, a_1, a_2, a_3 adalah konstanta variabel independen hasil regresi, dan e adalah jumlah error

D. Error

Error adalah jarak antara data yang dihasilkan antara titik data asli dengan titik data regresi. Error yang besar pada sebuah titik menandakan bahwa fungsi regresi tidak bagus fitnya pada titik tersebut. Secara matematis, error bisa didefinisikan sebagai berikut

$$e = y - f(x) \quad (2)$$

Dimana y adalah titik data asli dan $f(x)$ adalah titik yang didapat dari fungsi regresi.

E. Residual Sum of Square

Residual Sum of Square (selanjutnya disebut sebagai RSS) pada dasarnya menentukan seberapa baik fungsi regresi bisa merepresentasikan data awal. RSS mengukur jumlah kesalahan yang tersisa antara fungsi regresi dan kumpulan data asli. Angka RSS yang lebih kecil menunjukkan fungsi regresi yang lebih sesuai dengan data.

Sesuai namanya, RSS mengukur kesesuaian regresi dengan menjumlahkan semua perbedaan antara data asli dengan data regresi lalu mengkuadratkannya. Secara matematis, RSS bisa didefinisikan sebagai berikut.

$$S_r = \sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - a_0 - a_1x - a_2x^2 - a_3x^3)^2 \quad (3)$$

F. Standard Error of Estimate

Standard Error of Estimate (selanjutnya disebut sebagai SEE) pada dasarnya adalah jarak rata-rata perbedaan antara titik data asli dengan titik fungsi regresi. Mirip seperti RSS, SEE juga dapat digunakan untuk menilai seberapa sesuai fungsi regresi terhadap data asli. Secara matematis, SEE dapat didefinisikan sebagai berikut

$$S_{y/x} = \sqrt{\frac{S_r}{n - (m + 1)}} \quad (4)$$

Di mana n adalah jumlah data dan m adalah orde fungsi regresi.

III. DATA YANG DIGUNAKAN

Studi kasus ini akan menggunakan data harga mobil bekas dengan model spesifik, disebut sebagai model A. Mobil ini memiliki harga jual baru sebesar Rp. 400 juta, dan akan terus turun seiring berjalannya waktu. Studi kasus ini memiliki beberapa asumsi terhadap data harga jual. Pertama, data berasumsi bahwa harga mobil tetap sebesar Rp. 400 juta baik dari tahun paling awal data sampai paling akhir. Selain itu, data juga berasumsi bahwa tidak ada inflasi yang menyebabkan daya beli berkurang dan harga mobil bertambah.

Data memiliki ukuran 25 baris dan 2 kolom. Kolom tahun produksi akan dijadikan sebagai variabel independent sedangkan kolom harga jual dijadikan sebagai variabel dependent. Keseluruhan data adalah sebagai berikut:

Tahun Produksi (X)	Harga Jual (Y)
2010	122000000
2010	117000000
2011	118000000
2014	120000000
2014	130000000
2015	135000000
2016	165000000
2016	140000000
2017	165000000
2017	140000000
2017	150000000
2019	185000000
2019	180000000

2020	190000000
2020	175000000
2021	200000000
2021	230000000
2021	190000000
2021	220000000
2022	280000000
2022	290000000
2022	275000000
2023	320000000
2024	360000000
2025	400000000

Table 1. Data Harga Jual Mobil

IV. METODE YANG DIGUNAKAN

Pada studi kasus ini, perhitungan fungsi regresi yang akan digunakan untuk memprediksi harga jual mobil adalah polynomial regression. Polynomial regression akan mengambil titik data x dan y dan mengeluarkan hasil berupa fungsi polynomial yang merupakan bentuk best-fit terhadap data aslinya. Secara matematis, metode ini bekerja menggunakan matrix dalam bentuk

$$Ax = B \quad (5)$$

Di mana,

$$A = \begin{bmatrix} \sum_{i=0}^n x_i^6 & \sum_{i=0}^n x_i^5 & \sum_{i=0}^n x_i^4 & \sum_{i=0}^n x_i^3 \\ \sum_{i=0}^n x_i^5 & \sum_{i=0}^n x_i^4 & \sum_{i=0}^n x_i^3 & \sum_{i=0}^n x_i^2 \\ \sum_{i=0}^n x_i^4 & \sum_{i=0}^n x_i^3 & \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i \\ \sum_{i=0}^n x_i^3 & \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i & n \end{bmatrix},$$

$$x = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix},$$

$$B = \begin{bmatrix} \sum_{i=0}^n y_i x_i^3 \\ \sum_{i=0}^n y_i x_i^2 \\ \sum_{i=0}^n y_i x_i \\ \sum_{i=0}^n y_i \end{bmatrix}$$

Salah satu kegunaan utama dari regresi adalah

Karena variabel yang ingin dicari adalah matrix x , maka perumusan matrix bisa diubah menjadi sebagai berikut

$$x = A^{-1}B \quad (6)$$

Di mana A^{-1} adalah matrix invers dari matrix A . Hasil dari metode ini akan menghasilkan matrix x yang beranggotakan a_0, a_1, a_2, a_3 , yaitu koefisien dari persamaan regresi yang berbentuk seperti persamaan (1).

Setelah persamaan regresi didapatkan, studi kasus juga akan menganalisa error yang dihasilkan dari fungsi regresi. Nilai error akan menjadi penentu seberapa baik fungsi regresi bisa memprediksi data aslinya. Analisa error akan dilakukan menggunakan SEE, yang sudah dirumuskan oleh persamaan (4).

Perhitungan hasil regresi serta error akan dilakukan secara otomatis menggunakan program yang ditulis dalam C. Program akan menampilkan hasil akhir di layar melalui console. Program yang dilampirkan pada link Github pada awalnya ditulis oleh Henry Forson, dimana terdapat beberapa modifikasi untuk menambahkan fitur seperti perhitungan SEE.

V. DISKUSI DAN ANALISA HASIL

```
Angka koefisien fungsi regresi
a0 = -1174265545388150.000000
a1 = 1750133801493.889900
a2 = -869471816.917400
a3 = 143985.624900

Fungsi Regresi adalah:
f(x)=(143985.624900 * x^3) + (-869471816.917400 * x^2) +
(1750133801493.889900 * x) + -1174265545388150.000000

Standard Error of Estimate adalah:
SEE=17045142.876
atau
SEE=17.045 juta
```

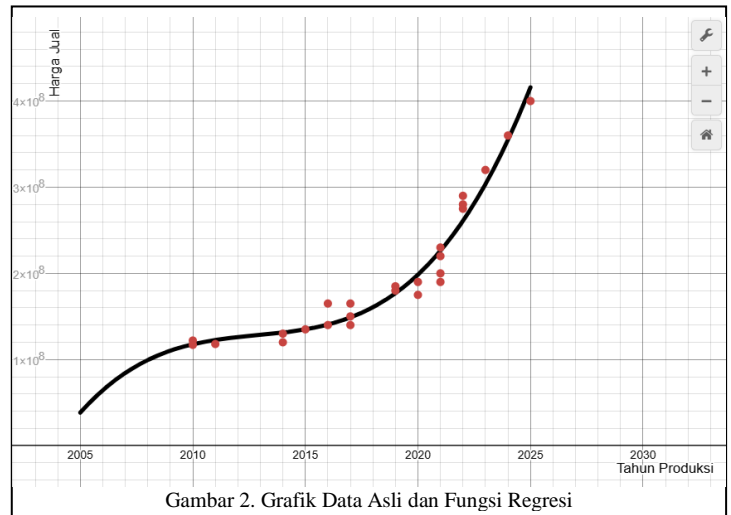
Gambar 1. Hasil Luaran Program

Program yang telah dijalankan memiliki 2 luaran utama. Pertama, program menampilkan koefisien fungsi serta fungsi regresi polinomial secara keseluruhan. Kedua, program menampilkan SEE dari regresi tersebut. Bila data pada tabel 1 dimasukkan ke dalam program, maka hasil luarannya akan sesuai dengan gambar 1. Secara lebih spesifik, luaran yang didapatkan dari program adalah sebagai berikut

$$f(x) = 143985.6259x^3 - 869471816.9174x^2 + 1750133801493.89x - 1174265545388150 \quad (7)$$

$$S_{y/x} = 17045142.876 \quad (8)$$

Bila dijabarkan dalam bentuk grafik, maka $f(x)$ serta data asli adalah seperti yang ada pada gambar (2). Dari luaran tersebut, dapat dilihat juga bahwa SEE dari hasil regresi sebesar Rp. 17.045 juta. Artinya, fungsi regresi dapat digunakan dengan keyakinan bahwa untuk sebagian besar kasus, data asli bisa berada sekitar \pm Rp. 17.045 juta dari hasil yang dikeluarkan oleh fungsi. Angka ini sudah cukup memadai, mengingat harga jual mobil yang sampai ratusan juta.



memungkinkan prediksi data menggunakan fungsi yang didapatkan. Oleh karena itu, studi kasus ini juga akan menampilkan prediksi berapa harga jual mobil yang sama bila tahun produksinya berada di kisaran tahun 2005 sampai dengan tahun 2009.

Tahun Produksi	Harga Jual
2005	Rp 38.340 \pm 17.045 juta
2006	Rp 64.336 \pm 17.045 juta
2007	Rp 84.400 \pm 17.045 juta
2008	Rp 99.395 \pm 17.045 juta
2009	Rp 110.19 \pm 17.045 juta

Tabel 2. Prediksi Harga Mobil Tahun Produksi 2005-2009

Melalui tabel (2), dapat dilihat harga yang bisa diharapkan bila ingin membeli atau menjual mobil dengan tahun produksi dari tahun 2005 sampai 2009.

VI. KESIMPULAN

Dari studi kasus yang telah dijalankan, bisa didapatkan beberapa kesimpulan singkat mengenai metode least square regression, yaitu

- Terdapat 2 metode yang bisa digunakan untuk memodelkan data, yaitu regresi dan interpolasi.
- Least square regression adalah salah satu metode regresi dengan kegunaan yang mirip dengan metode interpolasi, yakni untuk memodelkan dan memprediksi data.
- Metode regresi cenderung lebih baik digunakan ketimbang interpolasi bila data yang didapatkan memiliki error margin sama sekali. Artinya, metode ini paling cocok digunakan untuk data yang didapatkan dari dunia nyata.
- Metode least square regression bisa berupa banyak bentuk, seperti linear, non-linear, dan polynomial.

LINK GITHUB

<https://github.com/Maxwell-Zefanya/tugproyuaskomnum>

LINK YOUTUBE

<https://www.youtube.com/watch?v=CO-3HVx2EjI>

REFERENSI

- [1] S. C. Chapra, and R. P. Cnale, "Numerical Methods for Engineers Seventh Edition" McGraw-Hill Education, 2 Penn Plaza, New York, chapter 6
- [2] The Investopedia Team, "Residual Sum of Squares (RSS): What It Is and How to Calculate It," Investopedia, Mar. 6, 2025. <https://www.investopedia.com/terms/r/residual-sum-of-squares.asp>
- [3] W. Valeria, "13.3 Standard Error of the Estimate," Open Library, 2022. <https://ecampusontario.pressbooks.pub/introstats/chapter/13-3-standard-error-of-the-estimate/>
- [4] M. F. Henry, "Polyfit," Github, Feb. 11, 2022. <https://github.com/henryfo/polyfit>