

Prediksi Harga Smartphone menggunakan Model Multiple Linear Regression, Ridge Regression, dan Lasso Regression

Dian Nurmawati

Prodi Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Surabaya
Surabaya, Indonesia
dian.22005@mhs.unesa.ac.id

Abdini Qolbi Sahlina

Prodi Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Surabaya
Surabaya, Indonesia
abdini.22015@mhs.unesa.ac.id

Muhammad Abduh

Prodi Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Surabaya
Surabaya, Indonesia
muhammadabduh.22032@mhs.unesa.ac.id

Elly Matul Imah

Prodi Sains Data, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Surabaya
Surabaya, Indonesia
ellymatul@unesa.ac.id

Abstract— Dalam penelitian ini bertujuan untuk melakukan prediksi harga smartphone dengan menggunakan algoritma Multiple Linear Regression (MLR), Ridge Regression, dan Lasso Regression. Metode ini melibatkan analisis terhadap berbagai spesifikasi smartphone, seperti merek, kapasitas penyimpanan, intensitas kamera, dan faktor lainnya. Data spesifikasi dikombinasikan dengan algoritma prediktif untuk meningkatkan ketepatan prediksi harga smartphone. Algoritma MLR, Ridge Regression, dan Lasso Regression dievaluasi menggunakan metrik seperti R Square, Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Mean Absolute Percentage Error (MAPE). Pengujian dilakukan pada dataset yang terdiri dari 1512 baris dan 22 kolom. Hasil penelitian menunjukkan bahwa model MLR memiliki performa terbaik dengan R Square sebesar 1,00, MSE yang sangat rendah (0,00), MAE yang mendekati nol, dan MAPE sekitar 4,13%. Model ini dapat digunakan untuk memprediksi harga smartphone dengan akurasi tinggi.

Keywords— *Prediksi, Harga, Smartphone, MLR, Lasso Regression, Ridge Regression, R Square, MSE, MAE, MAPE*

I. PENDAHULUAN

Harga adalah suatu nilai tukar, dimana digunakan dalam pemasaran dan bisnis yang paling efektif. Harga menjadikan patokan pertama dalam pembelian dan penjualan. Setiap pelanggan pada saat ingin membeli sebuah ponsel pasti memikirkan spesifikasi ponsel tersebut yang akan dibelinya sesuai dengan estimasi harga yang disiapkan. Prediksi harga smartphone merupakan aspek penting dalam mengantisipasi dinamika pasar teknologi yang terus berubah. Dengan mempertimbangkan faktor-faktor seperti inovasi teknologi, persaingan industri, dan permintaan konsumen, perusahaan dapat menggunakan analisis prediktif untuk meramalkan tren harga smartphone di masa depan. Prediksi ini tidak hanya membantu perusahaan dalam perencanaan stok dan produksi, tetapi juga memungkinkan mereka untuk menyesuaikan strategi pemasaran dan penentuan harga agar lebih responsif terhadap kebutuhan pasar. Dengan demikian, prediksi harga smartphone tidak hanya menjadi alat

untuk mengoptimalkan keuntungan perusahaan tetapi juga untuk memenuhi ekspektasi konsumen dalam lingkungan yang terus berkembang secara teknologi [1].

Dalam penelitian ini, dilakukan uji prediksi harga smartphone dengan mempertimbangkan berbagai spesifikasi, seperti merek, kapasitas penyimpanan, intensitas kamera, dan faktor lainnya. Metodenya melibatkan penggunaan algoritma khusus untuk mencari nilai prediksi harga yang akurat. Pendekatan ini memungkinkan analisis lebih mendalam terhadap hubungan antara fitur-fitur teknis dan nilai pasar smartphone. Dengan menggabungkan data spesifikasi dengan algoritma prediktif, penelitian ini bertujuan untuk meningkatkan ketepatan prediksi harga smartphone, memberikan wawasan yang berharga bagi industri teknologi dan konsumen dalam pengambilan keputusan terkait pembelian smartphone [2].

Algoritma yang digunakan yaitu algoritma Model Multiple Linear Regression, Ridge Regression, dan Lasso Regression. Permasalahan yang mungkin muncul dalam analisis untuk membandingkan 3 model ini dengan menggunakan multiple linear regression, Ridge Regression, dan Lasso Regression adalah potensi di antara variabel independennya. Oleh karena itu, penelitian ini menekankan pentingnya evaluasi model untuk memastikan keakuratan analisis. Dalam konteks ini, R Square digunakan sebagai indikator uji korelasi antara variabel dependen dan variabel independen. Sementara itu, evaluasi dilakukan dengan mempertimbangkan Mean Absolute Error (MAE) dan Mean Absolute Percentage Error (MAPE) untuk mengukur sejauh mana prediksi model sesuai dengan data aktual atau harga. Dengan memanfaatkan metode evaluasi ini, penelitian bertujuan untuk memperoleh pemahaman yang lebih mendalam tentang kinerja model regresi dan memastikan hasil analisis yang lebih akurat dan dapat diandalkan.[2].

II. KAJIAN LITERATUR

A. Regresi

Regresi adalah metode statistik yang digunakan untuk memperkirakan hubungan antara satu atau lebih variabel independen (faktor penjelas) dan variabel dependen (hasil yang diinginkan). Rumus regresi linier sederhana adalah salah satu metode regresi yang digunakan untuk menentukan hubungan antara satu variabel dependen dan satu variabel independen. Berikut adalah rumus regresi linier sederhana:

$$Y = a + bX$$

di mana:

Y = variabel kriterium (hasil yang diinginkan)

a = variabel konstan

b = koefisien regresi

X = variabel independen (penjelas)

B. Smartphone

Smartphone merupakan salah satu jenis telepon seluler yang memiliki beragam fungsi selain hanya sebagai alat komunikasi suara. Alat ini menggabungkan berbagai fitur ke dalam satu perangkat, menjadikannya serbaguna dan cocok untuk berbagai tugas. Dengan sistem operasi (OS) yang ada pada smartphone, banyak berbagai fitur yang ada di dalam smartphone yang mendukung kebutuhan manusia pada umumnya dan hampir mendekati fungsi komputer[3]. Smartphone telah menjadi bagian penting dalam kehidupan sehari-hari banyak orang, menggantikan fungsi beberapa perangkat terpisah dengan satu perangkat ringkas dan multifungsi.

C. Prediksi

Prediksi adalah suatu upaya atau proses untuk memperkirakan atau menyatakan sesuatu yang akan terjadi di masa depan berdasarkan informasi, data, atau pengetahuan yang tersedia saat ini. Prediksi dapat berhubungan dengan berbagai bidang seperti ilmu pengetahuan, bisnis, olah raga, cuaca, dan lain-lain.

D. Harga

Harga adalah nilai moneter atau nilai tukar yang ditetapkan atau diberikan untuk suatu barang atau jasa. Harga merupakan salah satu elemen utama sistem ekonomi pasar, dan proses penentuannya dapat dipengaruhi oleh berbagai faktor, seperti penawaran dan permintaan, biaya produksi, persaingan pasar, dan kebijakan pemerintah.

E. Matriks

Matriks adalah kumpulan bilangan atau unsur yang tersusun dalam baris dan kolom. Baris dan kolom tersebut diisi oleh data-data yang disebut elemen matriks[4]. Matriks digunakan untuk mewakili dan memanipulasi data dalam berbagai bidang, termasuk matematika, ilmu komputer, fisika, ekonomi, dan statistik.

F. R-Squared

R-Squared mempunyai nilai antara 0 sampai 1 dan dapat diartikan sebagai persentase variasi variabel dependen yang dapat dijelaskan oleh model. Nilai R-squared yang lebih

tinggi menunjukkan bahwa model tersebut menjelaskan variasi data dengan lebih baik. Semakin besar R^2 maka semakin baik model regresinya. Untuk mencari nilai R squared, sebagaimana dalam bentuk persamaan seperti dibawah ini[5]:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Dimana:

R^2 : R squared

SSres : Jumlah kuadrat regresi

SStot : Jumlah total kolom

G. MAE (Mean Absolut Error)

MAE mengukur rata-rata dari nilai absolut dari selisih antara nilai prediksi dan nilai observasi aktual. Untuk mencari nilai MAE, sebagaimana dalam bentuk persamaan seperti dibawah ini:

$$MAE = \sum_{t=1}^n \frac{|Y' - Y|}{n}$$

Dimana:

MAE : Mean absolute error

Y' : Data prediksi

Y : Data aktual

n : Banyaknya data

H. MSE (Mean Squared Error)

MSE mengukur rata-rata dari nilai kuadrat dari selisih antara nilai prediksi dan nilai observasi aktual. Untuk mencari nilai MSE, sebagaimana dalam bentuk persamaan dibawah ini:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Dimana:

MSE : Mean Squared Error

n : jumlah pengamatan

y_i : nilai observasi aktual

\hat{y}_i : nilai prediksi.

I. MAPE (Mean Absolut Percentage Error)

MAPE mengukur rata-rata dari persentase nilai absolut dari selisih antara nilai prediksi dan nilai observasi aktual. Untuk mencari nilai MAPE, sebagaimana dalam bentuk persamaan seperti dibawah ini:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y - Y'|}{Y} \times 100\%$$

Dimana:

MAPE : Mean absolute percentage error

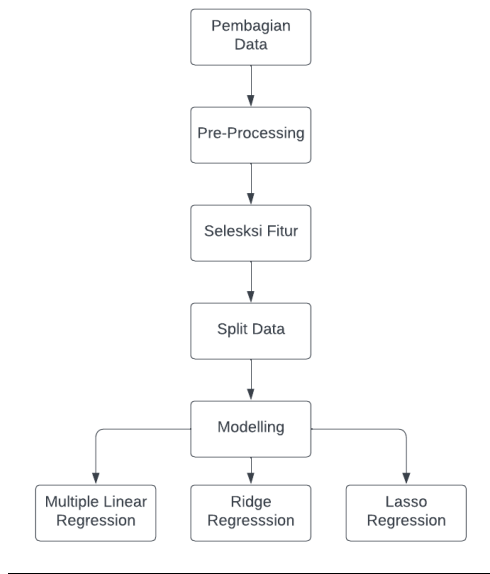
Y' : Data prediksi

Y : Data aktual

n : Banyaknya data

III. METODE

Gambar 1
Diagram Alir Metode



A. Pengambilan Data

Pada paper ini proses pengambilan data yang digunakan untuk prediksi harga smartphone didapat dari kaggle [3] dengan jumlah 1512 baris dan 22 kolom. Dataset ini berisi spesifikasi dari ponsel. Pada dataset ini terdapat dua jenis variabel yang digunakan yaitu 21 variabel prediksi dan satu variabel respon. Adapun 21 variabel prediksi diantaranya phone name, brand, os, inches, resolution, battery, battery type, ram(GB), announcement date, weight(g), storage(GB), video 720p, video 1080p, video 4K, video 8K, video 30fps, video 60fps, video 120fps, video 240fps, video 480fps, video 960fps. Sedangkan variabel respons adalah price (USD). Namun, pada paper ini peneliti melakukan drop features sebanyak 2 pada variabel prediksi yakni phone name dan announcement date dikarenakan program tidak dapat mengolah data selain numerik.

B. Pre-processing

Tahap pre-processing bertujuan untuk mengubah data mentah menjadi data yang berkualitas sehingga data layak untuk diolah pada tahapan selanjutnya. Tahapan ini dilakukan pada data mentah untuk menghilangkan data yang bermasalah atau inkonsisten. Data yang bermasalah seperti data yang noise atau yang mengandung error [6]. Sehingga data menjadi lebih terstruktur dan siap digunakan untuk proses analisis lebih lanjut.

C. Check Isna Data

Pengecekan check isna termasuk dalam tahap pre-processing yang dimana bertujuan untuk mengidentifikasi kata-kata atau token yang mungkin menunjukkan nilai yang hilang atau tidak valid. Dengan tujuan untuk memastikan bahwa dataset yang digunakan dalam analisis atau pemodelan data adalah data yang konsisten, bersih, dan dapat diandalkan. Hal ini mendukung proses pengambilan keputusan yang lebih baik dan hasil yang lebih akurat.

D. Encoded Data

Label encoding mengacu pada proses transformasi label kata menjadi bentuk numerik maupun bentuk variabel lain, dalam hal regresi (MLR) jika memuat variabel kategori dan nilainya tidak bisa di faktorisasi dalam bentuk tingkatan, dilakukan proses dummy, setiap nilai dalam variabel itu menjadi variabel lain. [7]

Label encoding pada penelitian ini mengubah semua nilai-nilai kategorikal dalam beberapa fitur DataFrame menjadi representasi numerik menggunakan LabelEncoder agar bisa dibaca oleh program, kemudian menyimpan DataFrame yang sudah di-encode ke dalam file CSV. Beberapa fitur yang diubah menjadi label encode diantaranya: brand, os, battery type, video 720p, video 1080p, video 4K, video 8K, video 30fps, video 60fps, video 120fps, video 240fps, video 480fps, video 960fps, dan resolution. Tujuan dari label encoding yaitu untuk mengubah data selain numerik agar dapat dimengerti atau dalam bentuk yang dapat dibaca oleh program.

E. Drop Column

Drop column atau menghapus kolom (fitur) yang tidak relevan atau tidak diperlukan bertujuan untuk membantu mengurangi dimensi data yang dapat bermanfaat dalam beberapa kasus, terutama jika terdapat banyak fitur (kolom) yang tidak memberikan kontribusi signifikan terhadap analisis atau model.

Pada penelitian ini penghapusan kolom (fitur) dilakukan pada fitur announcement date dan phone name. Penghapusan ini dilakukan karena pada kolom "phone name", terdapat informasi tipe dari setiap smartphone yang termasuk dalam nama brand. Oleh karena itu, untuk tujuan regresi, hanya informasi nama brand yang dianggap relevan, dan nama tipe dihapus. Sehingga, hanya kolom "brand" yang digunakan dalam analisis regresi.

F. Normalisasi

Normalisasi adalah pembentukan ulang data sehingga data yang mempunyai nilai ambiguity bisa dihilangkan. (Soeherman, 2004). Ada banyak jenis normalisasi yang biasa digunakan untuk membentuk ulang data.

Pada penelitian ini jenis normalisasi yang digunakan adalah Normalisasi Min-Max. Metode normalisasi min-max digunakan untuk mengatur ulang nilai data dengan nilai antara 0 sampai 1. (Chamidah, and Salamah, 2016). Metode normalisasi min-max dinyatakan sebagai berikut:

$$\text{normalized } x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Keterangan:

x = nilai asli data

x_{\min} = nilai minimal dari keseluruhan data

x_{\max} = nilai maksimal dari keseluruhan data

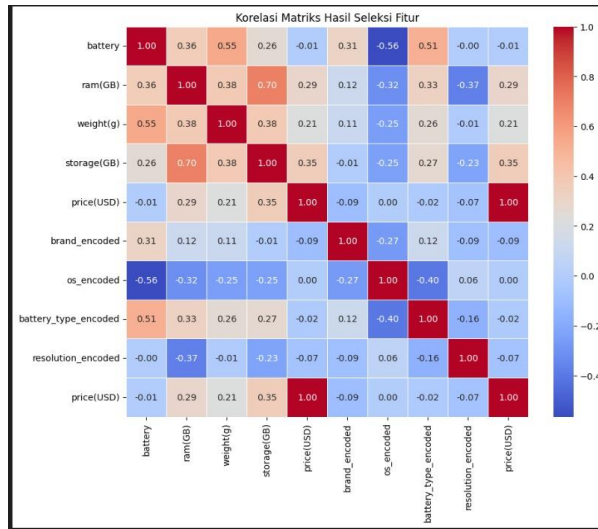
G. Seleksi Fitur

Seleksi fitur adalah cara untuk mengurangi jumlah fitur dan karenanya mengurangi kompleksitas komputasi model. Seringkali seleksi fitur menjadi sangat berguna untuk mengatasi masalah overfitting. Ini membantu kami dalam menentukan kumpulan fitur terkecil yang diperlukan untuk memprediksi variabel respons dengan akurasi tinggi.

Pada penelitian ini jumlah fitur yang digunakan adalah sebanyak 9 fitur untuk analisis regresi berganda, diantaranya: battery, ram(GB), weight(g), storage(GB),

price(USD), brand encoded, os encoded, battery type encoded, dan resolution encode. Berikut merupakan hasil korelasi matriks dari seleksi fitur:

Gambar 2
Korelasi Matriks Hasil Seleksi Fitur



H. Pembagian Data (Split Data)

Data-data yang sudah dilakukan pra proses data, kemudian data dipecah menjadi 2 jenis data yaitu data latih (training) dan data uji (testing). Data tersebut akan dilanjutkan ke proses pemodelan menggunakan MLR, Ridge Regression, dan Lasso Regression. Untuk tahapan data latih (training) jumlah sampel data latih harus lebih banyak standarnya 80% untuk data latih yang dipilih. Sedangkan sisanya data uji (testing) di dalam penelitian ini yaitu 20% data untuk diujikan. Tujuan dari data training pada penelitian ini yaitu membentuk sebuah pemodelan algoritma MLR dalam memprediksi harga sebuah smartphone. Sedangkan pada data testing hanya untuk menguji pemodelan pada algoritma MLR yang telah dibangun. Karena itu, data yang ada pada data testing seharusnya tidak boleh ada pada data training sehingga dapat diketahui apakah pemodelan sudah memenuhi target untuk mencapai hasil prediksi. Berikut merupakan tabel hasil dari split data:

Tabel 1
Hasil Split Data

X_train	(1209, 9)
X_test	(303, 9)
Y_train	(1209)
Y_test	(303)

I. Multiple Linier Regression

MLR atau regresi linier berganda juga merupakan regresi yang memiliki satu variable dependent (tidak bebas) dan lebih dari satu variable independen (bebas). Analisa ini untuk mengetahui arah hubungan antara variable, apakah masing-masing variable independen berhubungan positif atau negative dan untuk memprediksi nilai dari variable independen mengalami kenaikan atau penurunan [11].

Dikatakan regresi linier berganda jika ada lebih dari satu variabel. bebas atau variabel terikat. Sebaliknya, dikatakan regresi linier sederhana apabila hanya ada satu variabel bebas dan satu variabel terikatnya [12]. Berikut merupakan persamaan matematis dari *Multiple Linier Regression*:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

di mana:

Y = variabel dependen

X = variabel independen

β = koefisien regresi

ϵ = kesalahan acak

J. Ridge Regression

Regresi Ridge mirip dengan regresi LASSO. Tidak seperti Regresi Lasso, Regresi Ridge tidak mendorong koefisien tepat ke nol, melainkan mengecilkan mereka menuju nol. Ini mencegah overfitting dengan mengurangi pengaruh individu fitur dan mempromosikan mode yang lebih stabil dan berperilaku baik. Regresi Ridge sangat berguna ketika berhadapan dengan set data di mana semua fitur input dianggap relevan tetapi mungkin rentan terhadap multikolinieritas atau overfitting [13]. Berikut merupakan persamaan matematis dari *Ridge Regression* :

$$\hat{\beta}^R = (X'X + cI)^{-1}X'y$$

di mana:

$\hat{\beta}^R$ = Estimasi koefisien regresi ridge

X = Matriks prediktor

y = Vektor respons

c = Konstanta bias

Ridge regression merupakan salah satu metode yang telah diusulkan untuk memperbaiki multikolinieritas dengan memodifikasi metode kuadrat terkecil.

K. Lasso Regression

Regresi Lasso adalah varian dari regresi linier yang memperluas model dengan memperkenalkan regularisasi L1 untuk mencegah overfitting dengan memilih subset fitur yang relevan dan mendorong koefisien fitur yang kurang penting menuju nol. Regresi Lasso sangat berguna ketika berhadapan dengan dataset berdimensi tinggi atau dataset di mana beberapa fitur mungkin tidak berkontribusi secara signifikan terhadap prediksi (Melkumova & Shatskikh, 2017). Regresi Lasso dari pustaka scikit-learn digunakan untuk mengimplementasikan [14]. Berikut merupakan persamaan matematis dari *Lasso Regression*:

$$\hat{\beta}^{Lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

Dimana:

$\hat{\beta}^{Lasso}$ = Estimasi koefisien regresi Lasso

y = Vektor variabel respons

x = Matriks prediktor

β = Koefisien regresi

λ = Parameter penalti

L. Evaluasi

Melakukan pengujian terhadap model-model yang bertujuan untuk mendapatkan model yang paling akurat. Untuk mengetahui tingkat hasil prediksi pada pemodelan yang telah dibuat, akan dilakukan pengukuran performa dengan melihat nilai uji korelasi menggunakan R Squared sedangkan untuk tingkat pencapaian peramalan menggunakan MSE, MAE dan MAPE.

IV. PEMBAHASAN

- **R-Squared (R^2)**

R-Squared atau koefisien determinasi yaitu evaluasi model ini mengukur akurasi pengaruh hubungan antara variabel X (independen) terhadap variabel Y (dependen). R squared merupakan angka yang berkisar antara 0 sampai 1 yang mengindikasikan besarnya kombinasi variabel X secara bersama – sama mempengaruhi nilai variabel Y [15]. Berikut merupakan hasil untuk nilai R Square untuk semua model regresi:

Tabel 2
Hasil Evaluasi R-Square

Model Regresi	R^2
MLR	1.00
Ridge Regression	0.99
Lasso Regression	0.98

Semakin besar R^2 maka semakin baik model regresinya, seperti pada 4 kategori pengelompokan nilai R squared tabel dibawah ini. [16]

Tabel 3
Hipotesis R-Square

Nilai	Hipotesis
>0,75 – 0,99	Korelasi kuat/substansial
>0,50 – 0,75	Korelasi moderat/edang
>0,25 – 0,50	Korelasi cenderung lemah
<0,25	Korelasi sangat lemah

Berdasarkan hasil R Square pada masing-masing model regresi terlihat bahwa model Multiple Linier Regression memiliki hasil R Square yang paling besar diantara Ridge Regresi dan Lasso Regresi. Sesuai dengan tabel hipotesis nilai korelasi MLR sebesar 1.0 tergolong kedalam korelasi yang kuat/substansial.

- **MSE (Mean Squared Error)**

MSE digunakan untuk mengukur perbedaan kuadrat rata-rata antara nilai prediksi dan nilai aktual. Nilai MSE yang lebih kecil menunjukkan kecocokan yang lebih baik antara model dengan data. MSE mengukur rata-rata dari kuadrat perbedaan antara nilai yang diprediksi oleh model dan nilai aktual dari data. Berikut merupakan hasil untuk nilai MSE pada semua model regresi:

Tabel 4
Hasil Evaluasi MSE

Model Regresi	MSE
MLR	0,00
Ridge Regression	0.000000000000132
Lasso Regression	0.0002

Berdasarkan hasil MSE pada masing-masing model regresi terlihat bahwa model Multiple Linear Regression memiliki hasil MSE sebesar 0,00 yang sinkron terhadap hasil R Square sebelumnya yakni sebesar 1,00. MSE mengkuadratkan selisih antara nilai aktual dan nilai prediksi untuk setiap titik data, kemudian mengambil rata-rata dari nilai-nilai tersebut. Semakin rendah nilai MSE, semakin baik model regresi tersebut dalam memodelkan data.

- **MAE (Mean Absolut Error)**

MAE dalah dua diantara banyak metode untuk mengukur tingkat keakuratan suatu model peramalan. Nilai MAE merepresentasikan rata – rata kesalahan (error) rentang antara hasil peramalan dengan nilai sebenarnya [17]. Berikut merupakan hasil untuk nilai MAE pada semua model regresi:

Tabel 5
Hasil Evaluasi MAE

Model Regresi	MAE
MLR	0.00
Ridge Regression	0.00000075
Lasso Regression	0.00945

- **MAPE (Mean Absolut Percentage Error)**

Sedangkan MAPE adalah Pengukuran statistik tentang akurasi perkiraan (prediksi) pada metode peramalan. Metode ini memberikan informasi seberapa besar kesalahan peramalan dibandingkan dengan nilai sebenarnya dari series tersebut [18]. Berikut merupakan hasil untuk nilai MAPE pada semua model regresi:

Tabel 6
Hasil Evaluasi MAPE

Model Regresi	MAPE
MLR	4,13 %
Ridge Regression	1816308494.27%
Lasso Regression	22965019600173.81%

Nilai MAPE dapat diinterpretasikan atau ditafsirkan ke dalam 4 kategori yaitu seperti pada dibawah ini. [15]

Tabel 7
Hipotesis MAPE

Nilai MAPE	Hipotesis
<10%	“Sangat Baik”
10-20%	“Baik”
>20-50%	“Wajar”
>50%	“Tidak Akurat”

Berdasarkan hasil nilai MAPE pada masing-masing model regresi terlihat bahwa model Multiple Linear Regression memiliki hasil MAPE sebesar 0.04125412541261354 dalam bentuk persen adalah sekitar 4.13%. Sehingga hasil MAPE dari model MLR masuk kedalam kategori hipotesis “Sangat Baik”.

• SUMMARY MODEL

Ringkasan hasil yang diperoleh dari kinerja ketiga model regresi tersebut seperti yang ditunjukkan ada tabel berikut ini:

Tabel 8
Result Summary Model

Model Regresi	R ²	MSE	MAE	MAPE
MLR	1.00	0.00	0.00	4.13%
Ridge Regression	0.99	0.000000000000132	0.000000075	1816308494.27%
Lasso Regression	0.98	0.002	0.00945	22965019600173.81%

Pada Tabel 8 disajikan hasil tingkat akurasi dalam prediksi harga sebuah smartphone pada beberapa model regresi, diantaranya MLR, Ridge Regresi, dan Lasso Regresi. Terlihat bahwa akurasi terbaik diperoleh pada model MLR dengan uji korelasi menggunakan R squared menunjukkan sebesar 1,00 termasuk korelasi “Kuat” karena dalam skala >0,75 – 0,99 (hipotesis R squared), sedangkan nilai MSE (Mean Squared Error) sebesar 0.00, kemudian rata-rata dari selisih error (MAE) antara data aktual terhadap prediksi sebesar 0.00 dan pencapaian peramalan menggunakan MAPE dari data harga aktual terhadap harga prediksi dapat dikatakan “Sangat Baik” karena akurasi antara harga aktual terhadap harga prediksi tingkat pencapaiannya yaitu sebesar 4.13%. (hipotesis MAPE).

V. KESIMPULAN

Dalam penelitian prediksi harga smartphone menggunakan tiga model algoritma regresi yakni Multiple linier regression, Ridge Regression dan Lasso Regression. Masing-masing dari algoritma regresi yang dipakai memunculkan sebuah evaluasi model yaitu uji korelasi menggunakan R-Square dan menghitung pencapaian ramalan menggunakan MAE, MSE, dan MAPE.

Berdasarkan hasil summary model dapat disimpulkan bahwa tingkat akurasi terbaik diperoleh menggunakan *Multiple Linier Regression* dengan hasil uji korelasi R-Squared sebesar 1,0 termasuk korelasi yang kuat atau substansial karena dalam skala >0,75 - 0,99 (hipotesis R-Squared). Dalam rentang prediksi pengujian tingkat pencapaian peramalan, nilai rata-rata dari selisih error (MSE) sebesar 0.00, kemudian rata-rata dari selisih error (MAE) antara data aktual terhadap prediksi sebesar 0.00, dan untuk MAPE yaitu sebesar 4.13% yang dimana dalam tabel 7 Hipotesis MAPE masuk kedalam kategori “Kuat”.

REFERENCES

- [1] Toni Muhayat, J. S. (2022). Prediksi harga Smartphone menggunakan Algoritma Multiple Linear Regression. *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 506-525.
- [2] Tri Indarwati, T. I. (2018). PENGGUNAAN METODE LINEAR REGRESSION UNTUK PREDIKSI. *Jurnal TIKomSiN*, 1-6.
- [3] Shiri, M., & Michel, T. P. (2020). The Smartphone as a Pacifying Technology. *Journal of Consumer Research* (Vol. 47, No. 1, January 2020).
- [4] Ardiyanti, D. M., (2016). MATRIKS JURNAL PRAKTIKUM MATEMATIKA DASAR. LABORATORIUM MATEMATIKA DASAR JURUSAN MATEMATIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS JEMBER.
- [5] Akossou, A. Y. J., & Palm, R. (2013). Impact of Data Structure on the Estimators R-Square And Adjusted R-Square in Linear Regression. *International Journal of Mathematics and Computation* (Vol. 20; Issue No. 3; Year 2013
- [6] <https://www.kaggle.com/datasets/berkayeserr/phone-prices/code>
- [7] Effendi, M.R. 2015. “Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa Dengan Metode Naïve Bayes.” 101-106
- [8] Chandra, A. (2018). Beberapa Cara untuk Preprocessing Data dalam Machine Learning. <https://medium.com/warung-pintar/beberapa-cara-untuk-preprocessing-data-dalam-machinelearning-13cef4294614> (diakses pada 20 Desember 2023)
- [9] Soeherman, B., 2004. Normalisasi Sebagai Salah Satu Metode Untuk Membangun Database Dalam Era Revolusi Teknologi Sistem Informasi. *Journal of Management and Business*, 3(1)
- [10] Chamidah, N., ., W. and Salamah, U., 2016. Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi. *Jurnal Teknologi & Informasi ITS* Smart, 1(1), p.28.
- [11] D. Tampubolon and D. Sariurna, “Implementasi Regresi Linier Berganda Untuk Memprediksi Tingkat Penjualan Alat Kelistrikan,” vol. 3, no. 1, pp. 176–185, 2020.
- [12] Siti Aisyah Mahfuza, Zulfian Azmi, Guntur Syahputra "Data Mining Untuk Mengestimasi Angka Kemiskinan Di Sumatera Utara Menggunakan Metode Regesi Linier Berganda" Vol 4, No 6 (2021).
- [13] L. E. Melkumova and S. Y. Shatskikh, “Comparing Ridge and LASSO estimators for data analysis,” *Procedia Eng.*, vol. 201, pp. 746–755, 2017, doi: 10.1016/j.proeng.2017.09.615.
- [14] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011
- [15] Suryanto, A. A., & Asfan, M. (2019). PENERAPAN METODE MEAN ABSOLUTE ERROR (MEA) DALAM ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PRODUKSI PADI. *SAINTEKBU: Jurnal Sains dan Teknologi* (Volume : 11 No. 1 Februari 2019).
- [16] Hair, Jr., et. al. (2011). *Multivariate Data Analysis*. Fifth Edition. New Jersey: PrenticeHall.
- [17] Izzah, A., & Ratna, W. (2017). Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier. *KINETIK* (Vol. 2, No. 3, Agustus 2017, Hal. 141-150).
- [18] Lewis, C. D., (1982). *Industrial and Business Forecasting Methods*. London: Butterworths.