

Prediksi harga *Smartphone* menggunakan Algoritma *Multiple Linear Regression*

Toni Muhayat¹, Jayanta, S.Kom., M.Si.², Nurul Chamidah, S.Kom., M.Kom.³
Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Jawa Barat 12450
tonim@upnvj.ac.id¹, jayanta@upnvj.com², nurul.chamidah@upnvj.ac.id³

Abstrak. Ponsel atau telepon genggam adalah salah satu bukti produk kecanggihan teknologi yang semakin maju, karena pada fungsi umumnya. Ponsel adalah sebuah alat komunikasi jarak jauh. Namun kini ponsel bisa multifungsi dan mengubah ke segala aspek kehidupan manusia mulai dari baca berita, main game, pemotretan, dan lain sebagainya. Ponsel tersebut semakin canggih sehingga orang-orang menyebut istilah baru dengan sebutan "*Smartphone*". Lebih dari 50% orang-orang didunia memiliki sebuah *smartphone*, karena *smartphone* adalah salah satu kebutuhan utama di setiap masing-masing individu dikarenakan pemakaiannya yang cukup efisien. Melihat seberapa banyaknya konsumen yang menggunakan *smartphone*, maka dari itu pada penelitian ini bertujuan untuk memprediksi harga (variabel dependen) sebuah *smartphone*. Pengujian prediksi ini berdasarkan dari spesifikasi komponen-komponen (variabel independen) yang ada pada *smartphone*. Dari variabel independen tersebut, merupakan salah satu hal yang sangat diperhatikan calon pembeli *smartphone* dalam memilih produk *smartphone*. Teknik dalam memprediksi harga sebuah *smartphone* yaitu menggunakan algoritma *multiple linear regression* dalam konsep *machine learning*. Hasil yang didapatkan dalam penelitian ini adalah ditemukan bahwa uji korelasi antara variabel independen yang mempengaruhi variabel dependen (harga) dalam performa pada algoritma *Multiple Linear Regression (MLR)* yaitu dalam R squared 0,80 dapat dikatakan sebagai korelasi yang "Kuat". Sedangkan dalam pencapaian peramalan antara hasil prediksi dengan data harga adalah 23,9% dapat dikatakan sebagai pencapaian ramalan "Wajar".

Kata kunci: Prediksi, Smartphone, MLR, Harga, Dataset.

1 Pendahuluan

Perkembangan teknologi seluler yang semakin hari semakin pesat merupakan hal yang menguntungkan bagi distributor *smartphone* yang menawarkan berbagai macam inovasi. Tidak bisa dipungkiri bahwa perkembangan teknologi seluler ini sedikit banyak telah berpengaruh terhadap segala aspek di dalam kehidupan manusia baik dari segi sosial, ekonomi bahkan politik. Penelitian ini dilatarbelakangi oleh kondisi masyarakat Indonesia yang kurang paham dalam memilah harga *smartphone* dan seringkali merasa bingung jika dihadapkan dengan berbagai pilihan harga *smartphone* tersebut. Oleh karena itu terkadang mereka membutuhkan informasi mengenai *smartphone* yang akan dibeli baik itu melalui toko online maupun offline. Biasanya, mereka sebelum membeli sebuah *smartphone* mereka mencari informasi mengenai harga *smartphone* tersebut akan tetapi terkadang informasi yang didapatkan tidak sepenuhnya sesuai dengan kebutuhan oleh karena itu sangat dibutuhkan informasi yang bisa dengan mudah diakses dan tidak membingungkan calon pembeli dan membantu dalam mempermudah memilih *smartphone* yang paling tepat untuk dibeli. [1]

Ada beberapa *brand smartphone* yang populer di Indonesia, [2] *brand* yaitu suatu identitas terutama pada *smartphone* sebagai bentuk citra/image dari sebuah asosiasi yang muncul yang mengorientasikan didalam benak para calon pembeli *smartphone* ketika mengingat. Indikator yang mencirikan brand yaitu logo yang mudah diingat, mudah dikenali, yang terpercaya, populer, dan menyesuaikan perkembangan zaman. [3] Pada periode kuartal kedua (Q2) pada tahun 2021, Xiaomi menjadi pemimpin pasar *smartphone* di Indonesia menurut *Firma riset IDC* dan *Counterpoint* pada kuartal II-2021. Xiaomi berhasil mempertahankan pangsa pasarnya yang mencapai lebih dari 25%, dibandingkan dari kompetitor lain seperti Oppo, Vivo, Realme, dan Samsung. Posisi Xiaomi berhasil menggeser Oppo yang sebelumnya duduk di peringkat pertama di kuartal satu 2021 lalu. [4] Memasuki pertengahan 2022 ini, berdasarkan data *International Data Corporation (IDC) Quarterly Mobile Phone Tracker Q1*, Jumat, 17 Juni 2022. Samsung menduduki peringkat teratas *smartphone* dari segi pengiriman dengan pangsa pasar 22% pada periode awal 2022, disusul oleh Oppo dengan pangsa pasar 20,2%. Kemudian Vivo menduduki tempat ketiga dengan pangsa pasar 17,1%, sedangkan Xiaomi mendapatkan pangsa pasar sebesar 14,6%. Dan yang terakhir, Realme mendapatkan pangsa pasar 12,3%. [5] Pada kesimpulan diatas bahwa *brand* atau citra merek dalam suatu pengaruh harga *smartphone* itu cenderung relatif, karena peminatan pada calon pembeli *smartphone* pada tingkat citra merek diukur dari bagaimana orientasi *brand* tersebut terhadap calon pembeli *smartphone*.

Pada penelitian ini, yaitu melakukan sebuah pengujian prediksi harga *smartphone* dengan mempertimbangkan spesifikasi *smartphone* seperti merek, kapasitas penyimpanan, intensitas kamera dan lain sebagainya dengan menggunakan sebuah algoritma untuk mencari sebuah nilai prediksi harga. Algoritma yang digunakan yaitu algoritma *multiple linear regression*, yaitu salah satu metode statistika yang digunakan untuk mengetahui hubungan fungsional sebuah variabel tidak bebas (*dependent variable*) dengan dua atau lebih variabel bebas (*independent variable*). Adapun tujuan dari *multiple linear regression* ini adalah mengetahui seberapa besar pengaruh dari beberapa variabel independen terhadap variabel dependen dan juga dapat meramalkan nilai variabel tidak bebas apabila seluruh variabel independen sudah diketahui nilainya. Pada analisis *multiple linear regression* dengan banyak variabel independen, sering timbul masalah karena adanya hubungan antara dua atau lebih variabel independen. Variabel independen yang saling berkorelasi disebut multikolinearitas. Permasalahan yang terjadi pada *multiple linear regression* dapat mengakibatkan hasil analisis yang kurang akurat karena multikolinearitas di antara variabel independen nya, maka dibutuhkanlah evaluasi model dan pada penelitian ini evaluasi *R squared* dan *adjusted R squared* dijadikan sebagai perhitungan uji korelasi antara variabel dependen terhadap variabel independen nya dan evaluasi *MAE* dan *MAPE* sebagai pencapaian ramalan antara data aktual atau harga terhadap hasil prediksi.

2 Tinjauan Pustaka

2.1 Multiple Linear Regression

Multiple linear regression (MLR) yaitu model regresi linear berganda atau lebih yang dimana cara kerjanya melakukan semacam prediksi atau peramalan dalam berbentuk data. [6] Data yang dihasilkan dari pemodelan *MLR* tersebut adalah data pengaruhnya dari beberapa variabel independen terhadap variabel dependen sehingga terjadi tingkat akurasi dari kedua jenis variabel tersebut yang biasa disebut korelasi. Bentuk model persamaan *MLR* adalah sebagaimana dalam persamaan (1) di bawah ini.

$$Y' = b_0 + b_1X_1 + \dots + b_nX_n \quad (1)$$

Ket.

Y' = Variabel prediksi

X = Variabel prediktor

b = Koefisien

2.2 Smartphone

Smartphone yaitu alat telekomunikasi yang didesain secara cerdas. 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. [7]

2.3 Prediksi

Prediksi yaitu sebuah peramalan atau perkiraan mengenai sesuatu yang akan terjadi secara pola pikir dalam diri manusia. Prediksi akan terjadi jika didasari oleh informasi pada masa sebelum terjadi prediksi dan tujuan dari prediksi yaitu melakukan tindakan yang diperlukan untuk bersiap menghadapi sesuatu yang akan terjadi. [8]

2.4 Harga

Harga adalah jumlah uang telah disepakati oleh pembeli dan penjual untuk ditukar dengan barang atau jasa dalam transaksi bisnis normal. Dengan kata lain, harga adalah nilai suatu barang yang ditentukan oleh penjual. [9] Ada juga yang mengatakan definisi harga adalah sejumlah uang yang dibebankan kepada konsumen untuk mendapatkan manfaat dari suatu produk barang atau jasa yang dibeli dari penjual atau produsen.

2.5 Matriks

Matriks yaitu sebuah bilangan dalam suatu tempat (kurung siku) membentuk sebuah persegi yang dimana terdapat jumlah baris dan kolom. Baris dan kolom tersebut diisi oleh data-data yang disebut elemen matriks. [10] Di dalam penelitian ini, pada nantinya akan membentuk korelasi matriks, sehingga akan membentuk beberapa jenis matriks yaitu matriks independen dan matriks dependen.

2.6 R Squared Dan Adjusted R Squared

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*. [11] Untuk mencari nilai *R squared*, sebagaimana dalam bentuk persamaan (2) seperti dibawah ini. [12]

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

Ket.

R^2 : R squared

SS_{res} : Jumlah kuadrat regresi

SS_{tot} : Jumlah total kolom

Semakin besar R^2 maka semakin baik model regresinya, seperti pada 4 kategori pengelompokan nilai *R squared* tabel dibawah ini. [11]

Tabel 1. Hipotesis R Square

Nilai	Hipotesis
>0,75 - 0,99	Korelasi yang kuat/substansial
>0,50 - 0,75	Korelasi yang moderat/sedang
>0,25 - 0,50	Korelasi cenderung lemah
<0,25	Korelasi sangat lemah

Perbedaan utama antara *adjusted R squared* dan *R squared* adalah bahwa *adjusted R squared* menggambarkan jumlah varian dari variabel *Y* yang diwakili oleh setiap variabel *X*, sementara *R squared* mengukur variasi yang hanya dijelaskan oleh variabel independen yang benar-benar mempengaruhi variabel dependen. Untuk mencari nilai *adjusted R squared*, sebagaimana dalam bentuk persamaan (3) seperti dibawah ini. [12]

$$R^2_{adj} = 1 - (1 - R^2) \frac{n-1}{n-p-1} \quad (3)$$

Ket.

R^2_{adj} : Adjusted R squared

R^2 : Hasil R squared

n : Banyaknya data

p : Banyaknya prediktor

2.7 MAE Dan MAPE

Mean Absolute Error (MAE) adalah 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. [13] Perhitungan *MAE* ini cukup sederhana, yaitu nilai mutlak antara nilai prediksi dengan nilai sebenarnya $|Y' - Y|$. Untuk mencari nilai *MAE*, sebagaimana dalam bentuk persamaan (4) seperti dibawah ini.

$$MAE = \sum_{t=1}^n \frac{|Y' - Y|}{n} \quad (4)$$

Ket.

MAE : Mean absolute error

Y' : Data prediksi

Y : Data aktual

n : Banyaknya data

Sedangkan *Mean Absolute Percentage Error (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. [14] Untuk mencari nilai *MAPE*, sebagaimana dalam bentuk persamaan (5) seperti dibawah ini.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y - Y'}{Y} \right| \times 100\% \quad (5)$$

Ket.

MAPE : Mean absolute percentage error

Y' : Data prediksi

Y : Data aktual

n : Banyaknya data

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

Tabel 2. Hipotesis MAPE

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

2.8 Data Mining

Data mining merupakan kombinasi beberapa cabang ilmu komputer yang merupakan sebuah proses menemukan pola pada suatu kumpulan data yang besar. Tujuan dari *data mining* yaitu mengambil intisari dari suatu pengetahuan untuk mengumpulkan data dan menghasilkan struktur data yang lebih mudah dimengerti. [16]

2.8.1 Data Cleaning

Data cleaning yaitu melakukan pembersihan data untuk memperkecil *noise* dan data yang tidak sesuai serta tidak relevan. Proses pembersihan ini melibatkan beberapa hal antara lain menghapus data duplikat, memeriksa ketidakkonsistenan data, dan memperbaiki kesalahan pada data. [16]

2.8.2 Dummy Variabel

Dummy variable adalah sebuah teknik pada atribut data yang bersifat data kualitatif yang harus dikuantitatifkan. Untuk mengkuantitatifkan atribut pada variabel kualitatif, maka dibentuklah variabel *dummy* dengan nilai 1 dan 0. Nilai 1 menunjukkan adanya, sedangkan nilai 0 menunjukkan tidak adanya ciri kualitas tersebut. Misalnya variabel jenis kelamin, jika nilai 1 digunakan untuk laki-laki maka nilai 0 menunjukkan bukan laki-laki (perempuan), atau sebaliknya. [17]

2.8.3 Label Encoding

Label encoding adalah sebuah teknik melakukan klasifikasi yang biasanya berurusan dengan banyak label. Label-label ini bisa dalam bentuk kata-kata, angka, ataupun sesuatu yang lain. *Label encoding* mengacu pada proses transformasi label kata menjadi bentuk numerik maupun bentuk variabel lain, dalam hal regresi

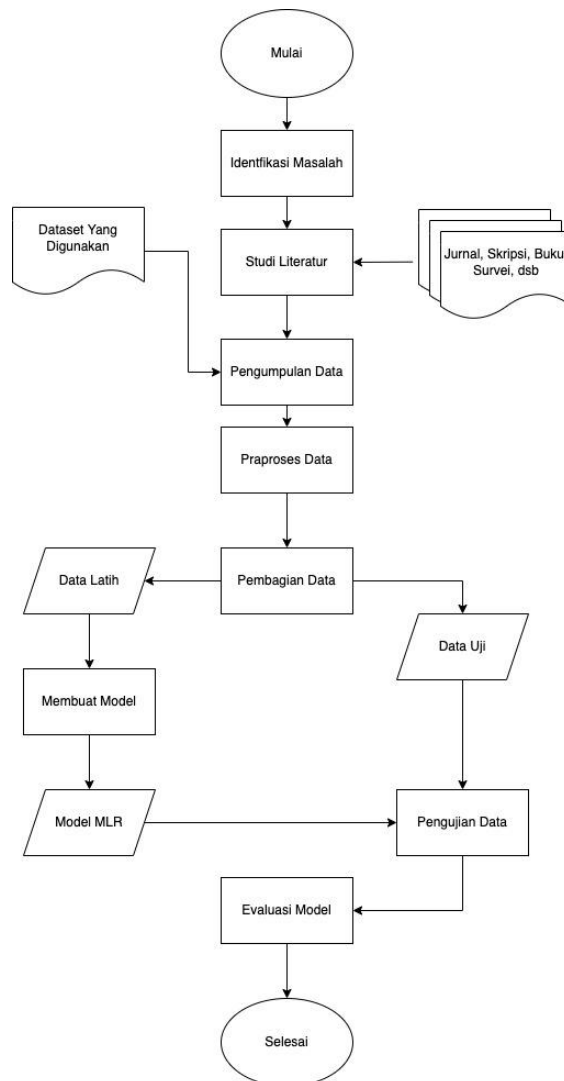
(*MLR*) jika memuat variabel kategori dan nilainya tidak bisa di faktorisasi dalam bentuk tingkatan, dilakukan proses dummy, setiap nilai dalam variabel itu menjadi variabel lain. [18]

2.9 Penelitian Terkait

Adapun jurnal penelitian yang relevan untuk dijadikan acuan untuk menunjang penelitian ini. Berikut ini paparan jurnal penelitian yang menjadi referensi yaitu:

- Penelitian dalam memprediksi harga saham yang dilakukan oleh Abidatul Izzah dan Ratna Widyastuti (2017), di dalam penelitiannya yaitu memprediksi harga saham menggunakan *MLR*. Pengujian diawali dengan pengumpulan data dan pra-proses data. Data harga saham yang akan digunakan dalam pengujian diperoleh dari laman *finance.yahoo.com* dengan kategori “*Jakarta Composite Index (^JKSE)*”. Selanjutnya proses prediksi dilakukan dengan *MLR* untuk mengatasi titik-titik saham yang outlier. Dari hasil yang diperoleh, dapat dilihat bahwa pendekatan paling baik ditunjukkan oleh metode *MLR* yakni dengan nilai *MAPE* sebesar 3,255. [14]
- Penelitian dalam memprediksi harga rumah yang dilakukan oleh Muhammad Labib Mu’tasim, dkk (2021), yaitu melakukan prediksi harga rumah dengan algoritma *MLR* berdasarkan dari beberapa faktor yaitu luas bangunan, luas halaman, jumlah kamar tidur, jumlah kamar mandi, dan domisili. Hasil dari penelitian ini memiliki tingkat akurasi 66% pada sebuah regresi antara harga rumah dengan kelima faktor tersebut. [19]
- Penelitian dalam memprediksi harga properti pada sebuah perusahaan yang dilakukan oleh Ghebyla Najla Ayuni dan Devi Fitriannah (2019), tujuan dari penelitian ini adalah untuk mendapatkan prediksi penjualan properti berdasarkan data penjualan properti pada PT XYZ. Berdasarkan hasil penelitian yang dilakukan dengan menggunakan data penjualan properti selama kurun waktu lima tahun, didapatkan bahwa jumlah penjualan pada bulan berikutnya untuk properti tipe rumah 59 diprediksi akan terjual sebanyak 4 unit, tipe rumah sudut sebanyak 2 unit, 6 tipe yang terdiri atas rumah 28, rumah 39, rumah 62, rumah 73, rumah 84, dan rumah 115 masing-masing sebanyak 1 unit, dan 5 tipe sisanya yaitu kavling, ruko, rumah 58, rumah 67, dan rumah 72 diprediksi tidak akan mengalami penjualan. Prediksi penjualan properti menggunakan metode regresi linear ini dapat dikatakan tergolong dalam kategori sangat baik didasari hasil pengujian keakuratan terhadap keseluruhan tipe yang menampilkan nilai *MAPE* yang memenuhi standar. [20]
- Penelitian dalam memprediksi penjualan *smartphone* yang dilakukan oleh Tri Irawati, dkk (2021), adalah untuk membuat sebuah sistem informasi yang dapat melakukan peramalan penjualan *smartphone* di toko 82 Cell Mayang dengan metode regresi linier. Sistem peramalan penjualan menggunakan metode tersebut, tujuannya untuk membuat sistem peramalan penjualan dengan menentukan volume penjualan dengan periode tertentu dengan melihat biaya iklan dan jumlah penjualan. Metode penelitian yang digunakan meliputi observasi, wawancara dan studi pustaka. Perancangan menggunakan DAD meliputi diagram relasi entitas, diagram konteks, hierarki proses input output dan diagram aliran data. Fitur-fitur pada aplikasi peramalan penjualan ini antara lain pengolahan data barang, data pelanggan, data produk yang masuk, data penjualan, dan data peramalan. Hasil pengujian menunjukkan nilai *MAPE* sebesar 0,032. Dari nilai tersebut dapat dikatakan bahwa prediksi penjualan *smartphone* dengan metode regresi linier pada 82 Cell Mayang dikategorikan sangat baik. [21]

3 Metode Penelitian



Gambar 1. Alur Penelitian

3.1 Identifikasi Masalah

Identifikasi masalah merupakan tahap untuk memperjelas masalah yang akan diangkat dalam penelitian ini yaitu permasalahan berapa prediksi harga sebuah *smartphone* yang dipasarkan berdasarkan dari kapasitas spesifikasi/fitur yang ada dengan metode algoritma *MLR* kemudian hasil prediksi dianalisa perbandingannya dengan harga pasar yang ada.

3.2 Perumusan Masalah

Rumusan masalah berdasarkan penelitian yang dibahas yaitu mengukur bagaimana penerapan algoritma *multiple linear regression* untuk memprediksikan harga sebuah *smartphone* dan sejauh mana performa pengujian model prediksi harga *smartphone* dengan menggunakan algoritma *multiple linear regression*.

3.3 Studi Literatur

Pada tahapan ini digunakan metode tinjauan pustaka yang dilakukan untuk pengumpulan ilmu pengetahuan dan informasi dari berbagai macam referensi antara lain skripsi, jurnal penelitian, buku, makalah, teori-teori serta sumber yang berasal dari internet seperti artikel ataupun video youtube pembelajaran dan lain sebagainya yang berkaitan dengan algoritma *MLR* ini, disertai dengan mata kuliah sebelumnya yang berkaitan dengan pembelajaran *data*

mining, machine learning, dan juga hasil survei lapangan ke setiap toko *smartphone* yang berkaitan dengan pertanyaan-pertanyaan yang sudah disusun mengenai pengaruh spesifikasi terhadap harga *smartphone*.

3.4 Pengumpulan Data

Penelitian ini menggunakan dataset yang didapatkan dari salah satu toko *smartphone* yang dikunjungi sebelumnya. Dataset tersebut, berisikan 6 brand dari 94 data pada bulan Desember tahun 2021. Data tersebut adalah data *smartphone* yang tersedia di toko. Dataset yang menjadi bahan penelitian ini terdiri dari 8 atribut yakni *Brand_type*, *RAM*, *ROM*, Jaringan, *Front_camera*, *Rear_camera*, dan *Battery_mAh* sebagai variabel independen. Sedangkan harga sebagai variabel dependen data yang ditampilkan belum dilakukan pra proses data sebelumnya.

3.5 Pra-Proses Data

Praproses data merupakan tahap awal yang dilakukan untuk mengolah data yang masih memiliki *noise* menjadi data yang siap untuk dilakukan pemodelan. Di dalam penelitian ini, ada 3 pra proses data sebelum melakukan pembentukan model *MLR* yaitu *data cleaning*, *dummy variabel*, dan *label encoding*.

3.5.1 Data Cleaning

Data cleaning yang perlu dilakukan pada atribut yang berisikan tipe data *string*. Contohnya pada atribut *Brand_Type* yang dimana nama merek terdapat nama tipe dari setiap *smartphone*, sehingga nama tipe perlu dihapus karena yang digunakan dalam regresi hanya dibutuhkan mana merek.

3.5.2 Dummy Variabel

Di dalam penelitian ini, hanya atribut *Brand_Type* yang hanya dilakukan *dummy variable* karena variabel bersifat kualitatif dan dalam atribut tersebut terdapat 6 jenis kategori, sehingga membentuk 6 jenis variabel *dummy* yang dimasukkan ke dalam pemodelan *MLR*.

3.5.3 Label Encoding

Label encoding pada penelitian ini mengubah semua nama atribut yang ada di dalam data menjadi beberapa nama variabel, seperti halnya variabel *D* sebagai variabel independen yang bersifat *dummy*, *X* sebagai variabel independen, dan *Y* sebagai variabel dependen. Tujuan dari *label encoding* yaitu untuk membuat data dapat dimengerti atau dalam bentuk yang dapat dibaca.

3.6 Pembagian 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*. 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.

3.7 Pemodelan

Penyelesaian prediksi harga *smartphone* dengan menggunakan algoritma *MLR* tersebut, ditangani dengan sistematis melalui proses penyelesaian dengan aturan korelasi matriks. Untuk membuat model pada algoritma tersebut digunakan 12 variabel independen dan 1 variabel dependen.

Bentuk persamaan *multiple linear regression* dalam penelitian ini sebanyak 12 bentuk persamaan variabel seperti pada persamaan (6) di bawah ini.

$$Y' = b_0 + b_1D_1 + \dots + b_6D_6 + b_7X_1 + \dots + b_{12}X_6 \quad (6)$$

Ket.

Y' : Hasil Prediksi

b : Koefisien

D : Variabel dummy

X : Variabel independen

Sebelum membuat pemodelan bentuk persamaan variabel, perlu diketahui bahwa untuk membuat pemodelan tersebut diperlukan beberapa matriks yang digunakan sebagai operasi korelasi pada setiap datanya. Ada 2 jenis matriks yang digunakan dalam korelasi yaitu matriks independen (X) dan matriks dependen (Y). Untuk setiap elemen pada matriks X dan Y adalah sebagaimana dalam persamaan (7) dibawah ini. [22]

Matriks X :

$$\begin{bmatrix} 1 & D1_1 & \dots & X5_1 & X6_1 \\ 1 & D1_2 & \dots & X5_2 & X6_2 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & D1_n & \dots & X5_n & X6_n \end{bmatrix}$$

(7)

Matriks Y :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}$$

Setelah dibentuk kedua jenis matriks, maka langkah selanjutnya yaitu mencari nilai koefisien (b) dari b0 sampai b12. Rumus persamaan untuk mencari nilai koefisien, tergantung dari elemen matriks yang di dapat, berikut ini adalah persamaan (8) untuk mencari nilai koefisien. [22]

$$b = (XTX)^{-1} \cdot XTY \quad (8)$$

Ket.

b : Koefisien

(XTX)-1 : Matriks invers dari hasil matriks X transpose dengan matriks X

XTY : Hasil perkalian matriks X transpose dengan Y

3.8 Pengujian Data

Pengujian data merupakan tahapan inti dari penelitian ini, pengujian data menggunakan 20% dataset yang digunakan untuk pengujian pada model *MLR* yang sudah dibuat dalam bentuk persamaan variabel serta melakukan input data baru memprediksi harga *smartphone* berdasarkan dari data inputan berupa spesifikasi-spesifikasi yang ada di dalam *smartphone*. Hasil dari prediksi pada data uji ini kemudian dilakukan evaluasi model baik itu evaluasi hasil prediksi maupun evaluasi kualitas model.

3.9 Evaluasi Model

Pada tahap ini merupakan tahapan evaluasi dari hasil pemodelan yang telah diterapkan. Untuk mengetahui tingkat hasil prediksi pada pemodelan yang telah dibuat, akan dilakukan pengukuran performa dengan melihat nilai uji korelasi menggunakan *R Squared* dan *Adjusted R squared*, sedangkan untuk tingkat pencapaian peramalan menggunakan *MAE* dan *MAPE*.

4 Hasil Dan Pembahasan

4.1 Data

Untuk mengawali penelitian ini dilakukan pengumpulan data dalam bentuk data harga *smartphone*. Data yang diperoleh meliputi *Brand_type*, *RAM*, *ROM*, *Jaringan*, *Front_camera*, *Rear_camera*, *Battery_mAh*, dan *Price_rupiah*. Terdapat 94 data yang tersedia dari beberapa *brand* ternama yang masuk pasar.

Tabel 3. Isi Dataset

No.	Brand_Type	RAM	ROM	Jaringan	Front_Camera	Rear_Camera	Battery_mAH	Price_rupiah
1	Realme Nazro 50a	4	64	4	8	50	6.000	2.199.000
2	Realme Nazro 50a	4	128	4	8	50	6.000	2.399.000
3	Realme Nazro 50i	4	64	4	5	8	5.000	1.649.000
...
93	Infinix Hot 9 Play	4	64	4	8	13	6.000	1.449.000
94	Infinix Smart 5	3	32	4	8	13	5.000	1.249.000

4.2 Pra Proses Data

4.2.1 Data Cleaning

Data cleaning digunakan untuk menghapus karakter yang tidak ada kaitannya dengan proses *MLR*, yaitu nama *type* pada setiap *brand*. Karena untuk *type* sendiri berfungsi sebagai penanda atau identitas dari spesifikasi setiap unit *smartphone*, sehingga yang diperlukan hanya nama *brand* saja yang mempengaruhi harga *smartphone*.

Tabel 4. Data Cleaning

No.	Brand_Type	RAM	ROM	Jaringan	Front_Camera	Rear_Camera	Battery_mAH	Price_rupiah
1	Realme	4	64	4	8	50	6.000	2.199.000
2	Realme	4	128	4	8	50	6.000	2.399.000
3	Realme	4	64	4	5	8	5.000	1.649.000

...
93	Infinix	4	64	4	8	13	6.000	1.449.000
94	Infinix	3	32	4	8	13	5.000	1.249.000

4.2.2 Dummy Variabel

Atribut yang harus dilakukan *dummy variabel* yaitu pada atribut *Brand_Type*, karena pada atribut tersebut bersifat data kualitatif (*string*) yang dimana data *string* tidak bisa melakukan pengolahan *MLR*. Maka dari itu dibentuklah sebuah kategori yang dimana pada nantinya akan membentuk notasi *dummy variabel*.

Tabel 5. Notasi Dummy Variabel

Kategori	D1	D2	D3	D4	D5	D6
Infinix	1	0	0	0	0	0
Oppo	0	1	0	0	0	0
Realme	0	0	1	0	0	0
Samsung	0	0	0	1	0	0
Vivo	0	0	0	0	1	0
Xiaomi	0	0	0	0	0	1

Ada 6 jenis kategori *dummy variabel* di dalam dataset yang digunakan, kemudian hasil *dummy* dari keseluruhan data.

Tabel 6. Dummy Variabel

No.	Infinix	Oppo	Realme	Samsung	Vivo	Xiaomi
1	0	0	1	0	0	0
2	0	0	1	0	0	0
3	0	0	1	0	0	0
...
93	1	0	0	0	0	0
94	1	0	0	0	0	0

4.2.3 Label Encoding

Label Encoding yaitu mengganti nama/label menjadi nama variabel, tujuannya supaya mudah untuk menentukan variabel dependen dengan independen. Pada variabel independen diberikan label *X* sebagai

data kuantitatif awal dan label *D* sebagai data *dummy* bersifat independen, sedangkan label *Y* sebagai variabel dependen.

Tabel 7. Label Encoding

No.	D1	D2	D3	D4	D5	D6	X1	X2	X3	X4	X5	X6	Y
1	0	0	1	0	0	0	4	64	4	8	50	6.000	2.199.000
2	0	0	1	0	0	0	4	128	4	8	50	6.000	2.399.000
3	0	0	1	0	0	0	4	64	4	5	8	5.000	1.649.000
...
93	1	0	0	0	0	0	4	64	4	8	13	6.000	1.449.000
94	1	0	0	0	0	0	3	32	4	8	13	5.000	1.249.000

4.3 Pembagian Data

Setelah pra proses data, maka tahap berikutnya yaitu data akan dibagi menjadi dua jenis data latih dan data uji. Sebelum melakukan pembagian data, data yang dilakukan terlebih dahulu melakukan inialisasi menjadi bagian variabel *X* independen dan variabel *Y* dependen. Selanjutnya setelah di inialisasi menjadi data *X* dan *Y*, kemudian dilakukan pembagian persentase data, di dalam penelitian ini 80% keseluruhan data dipecah untuk dijadikan data latih dan 20% sisanya untuk data uji.

Tabel 8. Persentase Data Latih dan Data Uji

Data Latih	Data Uji	Total Data
80%	20%	100%
75	19	94

Hasil pembagian data dilakukan secara random di dalam sebuah pemrograman, sehingga menghasilkan 80% data latih.

Tabel 9. Data Latih

No.	D1	D2	D3	D4	D5	D6	X1	X2	X3	X4	X5	X6	Y
1	0	0	0	1	0	0	3	32	4	5	15	5.000	1.499.000
2	0	1	0	0	0	0	8	128	4	32	48	4.310	4.099.000
3	0	0	0	1	0	0	6	128	4	8	57	5.000	2.499.000
...
74	0	0	0	0	0	1	12	256	5	16	108	5.000	7.499.000

No.	D1	D2	D3	D4	D5	D6	X1	X2	X3	X4	X5	X6	Y
75	0	0	0	1	0	0	4	64	4	5	17	5.000	1.799.000

4.4 Pemodelan

4.4.1 Matriks Independen (X)

Matriks independen yaitu matriks yang berisikan data-data independen termasuk data *dummy* yang ada didalam data latih. Matriks ini berukuran 75x13 dan setiap nilai elemen pada matriks ditentukan sebagaimana dalam bentuk persamaan (7). Dalam membuat regresi pada matriks, pada kolom elemen pertama diisi dengan angka 1 di setiap datanya hal ini karenakan untuk mencari nilai koefisien dari b_0 sebagai nilai intersep. Sehingga matriks tersebut menghasilkan elemen seperti pada dibawah ini.

$$\begin{bmatrix} 1 & 0 & \dots & 15 & 5.000 \\ 1 & 0 & \dots & 48 & 4.310 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 108 & 5.000 \\ 1 & 0 & \dots & 17 & 5.000 \end{bmatrix}$$

4.4.2 Matriks Dependen (Y)

Matriks dependen yaitu matriks yang berisikan data-data dependen atau data harga dari sebuah *smartphone*. Matriks ini berukuran berukuran 75x1 dan setiap nilai elemen pada matriks ditentukan sebagaimana dalam bentuk persamaan (7), sehingga menghasilkan elemen matriks seperti pada di bawah ini.

$$\begin{bmatrix} 1.499.000 \\ 4.099.000 \\ \dots \\ 7.499.000 \\ 1.799.000 \end{bmatrix}$$

4.4.3 Matriks Transpose X (X^T)

Setelah membentuk kedua jenis matriks, selanjutnya matriks independen kemudian di *transpose* kan menjadi matriks X^T berukuran 13x75. Sehingga menghasilkan bentuk *matriks* seperti pada dibawah ini.

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 15 & 48 & \cdots & 108 & 17 \\ 5.000 & 4.310 & \cdots & 5.000 & 5.000 \end{bmatrix}$$

4.4.4 Matriks XTX

Hasil dari *transpose* matriks X , kemudian dikalikan dengan matriks independen (X). Diketahui bahwa matriks XT berukuran 13×75 dan matriks X berukuran 75×13 , maka hasil perkalian tersebut membentuk matriks berukuran 13×13 pada matriks XTX . Sehingga menghasilkan bentuk matriks seperti dibawah ini.

$$\begin{bmatrix} 75 & 13 & \cdots & 3791 & 374.870 \\ 13 & 13 & \cdots & 484 & 71.300 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 3.791 & 484 & \cdots & 250.993 & 18.659.800 \\ 374.870 & 71.300 & \cdots & 18.659.800 & 1.889.708.900 \end{bmatrix}$$

4.4.5 Matriks Invers XTX / (XTX)⁻¹

Hasil dari matriks XTX kemudian di inverskan. Sehingga menghasilkan bentuk matriks seperti dibawah ini.

$$\begin{bmatrix} 1,38572E+15 & -1,38572E+15 & \cdots & -0,045408913 & 0,003235474 \\ -1,38572E+15 & 1,38572E+15 & \cdots & 0,04948387 & -0,004224741 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -0,006806551 & 0,010881508 & \cdots & 4,68843E-05 & -6,81803E-07 \\ -0,002090988 & 0,001101721 & \cdots & -6,81803E-07 & 1,65173E-07 \end{bmatrix}$$

4.4.6 Matriks XTY

Hasil dari *transpose* matriks X , kemudian dikalikan dengan matriks dependen (Y). Diketahui bahwa matriks XT berukuran 13×75 dan matriks Y berukuran 75×1 , maka hasil perkalian tersebut membentuk matriks berukuran 13×1 pada matriks XTY . Sehingga menghasilkan elemen matriks seperti dibawah ini.

$$\begin{bmatrix} 266.315.000 \\ 25.977.000 \\ 42.091.000 \\ 33.140.000 \\ 44.788.000 \\ 25.893.000 \\ 94.426.000 \\ 1.887.814.000 \\ 40.866.976.000 \\ 1.162.091.000 \\ 5.232.625.000 \\ 15.984.979.000 \\ 1,29518E+12 \end{bmatrix}$$

4.4.7 Matriks $(X^T X)^{-1}X^T Y$

Kemudian untuk mencari nilai koefisien (b) maka lakukan perkalian antara matriks *invers* $X^T X$ dengan matriks $X^T Y$ yang sudah dibentuk. Sehingga akan mengetahui nilai koefisien dari b_0 s/d b_{12} seperti pada matriks dibawah ini.

$$\begin{bmatrix} -4.262.649,689 \\ -761.250,2423 \\ 724.842,7959 \\ 15.725,11009 \\ -111.806,7863 \\ 156.106,8811 \\ -23.617,75846 \\ 347.338,2173 \end{bmatrix}$$

5.368,716237
894.657,4409
34.387,67005
6.687,052314
76,812741

Sehingga terbentuklah bentuk persamaan variabel atau rumus untuk memprediksi suatu harga pada sebuah *smartphone* (Y') menjadi seperti pada persamaan (12) dibawah ini.

$$Y' = (-4.262.649,689) + (-761.250,2423)(D1) + 724.842,7959(D2) + 15.725,11009(D3) + (-111.806,7863)(D4) + 156.106,8811(D5) + (-23.617,75846)(D6) + 347.338,2173(X1) + 5.368,716237(X2) + 894.657,4409(X3) + 34.387,67005(X4) + 6.687,052314(X5) + 76,812741(X6) \quad (12)$$

4.5 Pengujian Data

Pada pengujian data, yaitu melakukan sebuah penginputan data kedalam model *MLR* yang sudah dibentuk, hal ini bertujuan untuk menampilkan hasil prediksi harga berdasarkan data-data yang sudah diinputkan melalui program. Berikut ini adalah kedua sampel dari ke 19 data uji dalam memprediksi harga *smartphone* secara manual, misalkan ada dua merek *smartphone* yaitu Xiaomi (sampel 1) dan Oppo (sampel 2) yang memiliki spesifikasi seperti pada tabel dibawah ini.

Tabel 10. Input Data Uji

Sampel	D1	D2	D3	D4	D5	D6	RAM	ROM	Jaringan	Kamera Depan	Kamera Belakang	Baterai	Harga
1	0	0	0	0	0	1	4	64	4	8	50	5.000	2.199.000
2	0	1	0	0	0	0	4	64	4	16	13	5.000	2.499.000

Kedua sampel data inputan di atas, akan dimasukan ke dalam pemodelan *MLR* pada bentuk persamaan (12), seperti pada penyelesaian dibawah ini.

$$\text{Harga sampel 1} = (-4.262.649,689) + (-761.250,2423)(0) + 724.842,7959(0) + 15.725,11009(0) + (-111.806,7863)(0) + 156.106,8811(0) + (-23.617,75846)(1) + 347.338,2173(4) + 5.368,716237(64) + 894.657,4409(4) + 34.387,67005(8) + 6.687,052314(50) + 76,812741(5.000)$$

$$\text{Harga sampel 1} = 2.018.830,705$$

$$\text{Harga sampel 2} = (-4.262.649,689) + (-761.250,2423)(0) + 724.842,7959(1) + 15.725,11009(0) + (-111.806,7863)(1) + 156.106,8811(0) + (-23.617,75846)(0) + 347.338,2173(4) + 5.368,716237(64) + 894.657,4409(4) + 34.387,67005(16) + 6.687,052314(13) + 76,812741(5.000)$$

$$\text{Harga sampel 2} = 2.794.971,684$$

Pada hasil diatas dapat dilihat bahwa *smartphone* sampel 1 memiliki prediksi harga 2.018.830,70 juta memiliki jarak rentang 180.169,29 ribu terhadap harga aslinya yaitu 2.199.000 juta. Sedangkan pada *smartphone* merek sampel 2 memiliki prediksi harga 2.794.971,68 juta memiliki jarak rentang 295.971,68 ribu lebih murah terhadap harga aslinya yaitu 2.499.000 juta.

Proses pengujian model untuk keseluruhan data dari 20% data uji, menghasilkan nilai prediksi seperti pada tabel dibawah ini.

Tabel 11. Sample Data Uji

No.	D1	D2	D3	D4	D5	D6	X1	X2	X3	X4	X5	X6	Y	Y'
1	0	0	1	0	0	0	4	64	5	5	8	5.000	1.649.000	2.568.811,807
2	0	1	0	0	0	0	4	64	4	16	13	5.000	2.499.000	2.794.971,684
3	0	0	0	0	0	1	6	128	4	13	64	5.000	2.799.000	3.322.662,062
...
18	0	0	0	0	1	0	2	32	4	5	13	4.030	1.499.000	906.987,6863
19	0	0	0	0	0	1	8	128	4	20	64	5.160	3.399.000	4.270.342,225

4.6 Evaluasi Model

4.6.1 R Squared dan Adj R Squared

Selanjutnya hasil pemodelan yang sudah dibentuk, kemudian dilakukan uji korelasi menggunakan evaluasi model *R squared* dan *adjusted R squared*. Hal yang pertama kali dilakukan yaitu mencari nilai *Sum Square Residual (SSres)* dan *Sum Square Total (SStot)*. Untuk mencari kedua fungsi tersebut, maka diperlukan rumus untuk mencarinya sebagaimana dalam persamaan (10) dan (11) seperti pada dibawah ini. [23]

$$SSres = \sum_{i=1}^n (Y - Y')^2 \quad (10)$$

$$SStot = \sum_{i=1}^n (Y - \bar{Y})^2 \quad (11)$$

Ket.

SSres : Jumlah kuadrat regresi

SStot : Jumlah total kolom.

Y : Data aktual

Y' : Data prediksi

\bar{Y} : Rata-rata data aktual

Dik :

$$\bar{Y} = \frac{(1.649.000 + 2.499.000 + \dots + 1.499.000 + 3.399.000)}{19}$$

$$= 3.259.526,316$$

Tabel 12. Mencari SSres & SStot

No.	Y	Y'	(Y-Y') ²	(Y- \bar{Y}) ²
1	1.649.000	2.568.811,807	8,46054E+11	2,5938E+12
2	2.499.000	2.794.971,684	87.599.237.990	5,784E+11
3	2.799.000	3.322.662,062	2,74222E+11	2,12084E+11
....
18	1.499.000	906.987,6863	3,50479E+11	3,09945E+12
19	3.399.000	4.270.342,225	7,59237E+11	19.452.908.529

$$SS_{res} = (8,46054E+11 + 87.599.237.990 + \dots + 3,50479E+11 + 7,59237E+11)$$

$$= 1,78852E+13$$

$$SStot = (2,5938E+12 + 5,784E+11 + \dots + 3,09945E+12 + 19.452.908.529)$$

$$= 9,21379E+13$$

$$R^2 = 1 - \left(\frac{1,78852E+13}{9,21379E+13} \right)$$

$$= 0,80588655$$

Pada hasil diatas, menjelaskan bahwa *R squared* memiliki tingkat korelasi pengaruh *X* terhadap *Y* yaitu 0,80 yang dimana didalam tabel hipotesis *R squared* (tabel hipotesis *R Squared*) nilai >0,75 - 0,99 termasuk korelasi kuat dari pengaruh semua variabel *X* independen tersebut terhadap *Y* dependen.

Setelah menentukan nilai *R squared*, kemudian menentukan nilai dari *adjusted R squared* (*R²adj*).

$$R^2_{adj} = 1 - (1 - 0,80588655) \times \left(\frac{(19 - 1)}{(19 - 12 - 1)} \right)$$

$$= 0,41765966$$

Pada hasil diatas, menjelaskan bahwa *adjusted R squared* memiliki tingkat pengaruh berbobot 41% dari nilai 0,41 yang dihasilkan.

4.6.2 MAE dan MAPE

Dalam pencapaian peramalan yaitu dengan menggunakan evaluasi model *MAE* dan *MAPE*. *Mean Absolute Error* (*MAE*) berfungsi untuk menunjukkan nilai kesalahan rata-rata *error* antara nilai aktual (*Y*) terhadap prediksi (*Y'*). Sedangkan pada *Mean Absolute Percentage Error* (*MAPE*) yakni mengukur akurasi peramalan dari *Y* terhadap *Y'* dalam bentuk persen. Pada evaluasi model *MAE* dan *MAPE* tersebut, berfungsi untuk menunjukkan pencapaian peramalan pada hasil di dalam data uji.

Sebelum melakukan evaluasi model, hal yang utama dilakukan yaitu menghitung variabel *E* (*error*). Karena, kepastian *error* itu sangat terjadi ketika melakukan prediksi, dan hasil prediksi tidak akan sampai terakurasi 100% terhadap data aktual (*Y*). Untuk mencari nilai *error* yakni sebagaimana dalam persamaan (9). [13]

$$E = |Y - Y'| \tag{9}$$

Ket.

E : Error (residual)

Y : Data aktual

Y' : Data prediksi

Sehingga menghasilkan nilai-nilai operasi seperti dibawah ini.

$$E = |1.649.000 - 2.568.811,807|, |2.499.000 - 2.794.971,684|, |.....|, |1.499.000 - 906.987,6863|, |3.399.000 - 4.270.342,225|$$

$$= |919.811,8074|, |295.971,6844|, |.....|, |592.012,3137|, |871.342,225|$$

Tabel 13. Mencari Error

No.	Y	Y'	E
1.	1.649.000	2.568.811,807	919.811,8074
2.	2.499.000	2.794.971,684	295.971,6844
3.	2.799.000	3.322.662,062	523.662,0616
....
18.	1.499.000	906.987,6863	592.012,3137
19.	3.399.000	4.270.342,225	871.342,225

Untuk mencari pencapaian peramalan menggunakan *MAE* dan *MAPE*, tabel yang perlu ditunjukkan yaitu nilai *E* atau $|Y'-Y|$ dan nilai *E* dibagi *Y* harga kemudian dikali 100%.

Tabel 14. Mencari Nilai MAE dan MAPE

No.	Y	Y'	$ Y'-Y $ (E)	$ (Y'-Y)/Y \times 100\%$
1	1.649.000	2.568.811,807	919.811,8074	55,77997619
2	2.499.000	2.794.971,684	295.971,6844	11,84360482
3	2.799.000	3.322.662,062	523.662,0616	18,70889824
....
18	1.499.000	906.987,6863	592.012,3137	39,49381679%
19	3.399.000	4.270.342,225	871.342,225	25,63525228%
SUM	61.931.000	56.440.908,96	12.369.150,39	398,3249221%

$$\begin{aligned} MAE &= \frac{12.369.150,39}{19} \\ &= 699.419,0629 \\ MAPE &= \frac{398,3249221\%}{19} \\ &= 23,9002578\% \end{aligned}$$

Jadi kesimpulannya pada evaluasi model *MAE* memperoleh nilai akurasi 699.419,0629 dan pada *MAPE* di perlihatkan bahwa hasil *MAPE* memperoleh pencapaian peramalan yaitu 23,9% yang artinya dikategorikan sebagai akurasi “Wajar” berdasarkan pada hipotesis *MAPE*.

5 Kesimpulan dan Saran

Kesimpulan pada penelitian ini yaitu penerapan *algoritma multiple linear regression* untuk memprediksi harga sebuah *smartphone* yaitu mencari dataset, setelah itu kemudian data dilakukan pra proses data, kemudian melakukan pembagian data, setelah itu membuat pemodelan untuk mencari bentuk persamaan atau rumus prediksinya. Setelah mengetahui bentuk persamaannya, lakukanlah pengujian data dan tahap terakhir yaitu melakukan evaluasi model yaitu uji korelasi menggunakan *R squared* dan *adjusted R squared* dan menghitung pencapaian peramalan menggunakan *MAE* dan *MAPE*.

Tingkat akurasi dalam prediksi harga sebuah *smartphone* pada uji korelasi menggunakan *R squared* menunjukkan sebesar 0,80 termasuk korelasi “Kuat” karena dalam skala $>0,75 - 0,99$ (hipotesis *R squared*), sedangkan pada *adjusted R squared* menunjukkan sebesar 41%. Dalam rentang prediksi pengujian tingkat pencapaian peramalan, nilai rata-rata dari selisih *error (MAE)* antara data aktual terhadap prediksi sebesar 699.419,0629 dan pencapaian peramalan menggunakan *MAPE* dari data harga aktual terhadap harga prediksi dapat dikatakan “Wajar” karena akurasi antara harga aktual terhadap harga prediksi tingkat pencapaiannya yaitu 23,9% (hipotesis *MAPE*).

Sara dari penelitian ini adalah, dataset untuk merek-merek *smartphone* hanya ada 6 jenis *brand* yakni Samsung, Xiaomi, Oppo, Vivo, Realme, dan Infinix. Tidak ada nama *brand* lain selain keenam nama tersebut, contohnya iPhone ataupun merk lain, sayangnya toko *smartphone* tempat penelitian tersebut sedang tidak tersedia. Kemudian tambahkan atribut lain yang dapat membentuk lebih dari 6 variabel *X* yang relevan, supaya lebih bervariasi keakuratan antara data aktual (harga) dengan prediksi serta pada evaluasi model pencapaian ramalan masih bisa ditingkatkan ke kategori “Sangat Baik” apabila jumlah sampel pada data latih dan uji bisa lebih banyak yang digunakan.

Referensi

- [1] Wirawan, D. F. (2019). PEMILIHAN HANDPHONE DENGAN METODE SAW (SIMPLE ADDITIVE WEIGHTING). *Artikel Skripsi Universitas Nisantara PGRI Kediri*.
- [2] Vanessa, I., & Zainul, A. (2017). PENGARUH CITRA MEREK (*BRAND IMAGE*) DAN HARGA TERHADAP KEPUTUSAN PEMBELIAN KONSUMEN. *Jurnal Administrasi Bisnis (JAB) (Vol. 51 No. 1 Oktober 2017)*.
- [3] Wardhaniika, N. I. K., & Ignatia, M. H. (2021). PERPINDAHAN MEREK AKIBAT KETIDAKPUASAN KONSUMEN DALAM PEMILIHAN PRODUK SMARTPHONE. *JURNAL EKONOMI & BISNIS (VOL. 2, NO. 1, JUNI 2021: 21–31)*.
- [4] Asikin, M. N., & Rian, A. (2021). Lima Nama Merek Ponsel Teratas Q2 2021, Siapa Saja?. <https://www.jawapos.com/oto-dan-teknogadget/26/10/2021/> (diakses pada 21 Juni 2022).
- [5] Azizah, N. (2022). Pasar Ponsel Turun, Ini Lima Merek Ponsel Teratas yang Diminati di Indonesia. <https://www.republika.co.id/berita/rdkwcy463/> (diakses pada 21 Juni 2022).
- [6] Ariyanti, N. D., et al. (2018). PENERAPAN METODE REGRESI LINIER BERGANDA DALAM PRODUKSI PAKAIAN GUNA MENUNJANG SISTEM PENDUKUNG KEPUTUSAN. *Jurnal Rekayasa Teknologi Nusa Putra (Vol. 5, No. 1, September 2018 : Hal 34-39)*.

- [7] Shiri, M., & Michel, T. P. (2020). The Smartphone as a Pacifying Technology. *Journal of Consumer Research* (Vol. 47, No. 1, January 2020).
- [8] Shilimkar, G., Shivam, P. (2021). Disease Prediction Using Machine Learning. *IRE Journals* (Volume 3 Issue 10, April 2020).
- [9] Yuliana, I. S.Pd. M.Pd., & Jonson (2017). PENGARUH HARGA DAN PROMOSI TERHADAP KEPUTUSAN PEMBELIAN PADA PT.ZUVACE MONDY DI KOTA BATAM. *Jurnal Akuntansi dan Bisnis (Akmenbis) Tahun 2017 Akademi Akuntansi Permata Harapan* (Vol. 2, No. 7, 2017).
- [10] Ardiyanti, D. M., (2016). MATRIKS JURNAL PRAKTIKUM MATEMATIKA DASAR. *LABORATORIUM MATEMATIKA DASAR JURUSAN MATEMATIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS JEMBER*.
- [11] Hair, Jr., et. al. (2011). *Multivariate Data Analysis. Fifth Edition. New Jersey: PrenticeHall.*
- [12] 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).
- [13] 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).
- [14] 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).
- [15] Lewis, C. D., (1982). *Industrial and Business Forecasting Methods. London: Butterworths.*
- [16] Shafarindu, A. I., et al. (2021). Penerapan Algoritma *Naive Bayes* untuk Klasifikasi Tingkat Kebugaran Jasmani Berdasarkan Hasil Pengukuran pada Pegawai. *e-ISBN 978-623-93343-4-5 Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)* (Jakarta-Indonesia, 15 September 2021).
- [17] Junaidi, (2015). Regresi dengan Variabel Dummy. Fakultas Ekonomi dan Bisnis Universitas Jambi : 4 Januari 2015.
- [18] Chandra, A. (2018). Beberapa Cara untuk Preprocessing Data dalam Machine Learning. <https://medium.com/warung-pintar/beberapa-cara-untuk-preprocessing-data-dalam-machine-learning-13cef4294614> (diakses pada 17 Juli 2022).
- [19] Mu'tasim, M. L., et al. (2021). ANALISIS HARGA RUMAH SESUAI SPESIFIKASI MENGGUNAKAN MULTIPLE LINEAR REGRESSION. *Jurnal FIK Upnvj 2021*.
- [20] Ghebyla, N. A., & Devi, F. (2019). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika* (vol. 14 no. 2, Institut Teknologi Harapan Bangsa, Bandung).
- [21] Indarwati, T., Tri I., & Elistya, R. (2018). PENGGUNAAN METODE LINEAR REGRESSION UNTUK PREDIKSI PENJUALAN SMARTPHONE. *Jurnal TIKomSiN* (Vol. 6, No. 2, Oktober 2018).
- [22] Utama, C. (2009). DENGAN PENDEKATAN MATRIKS DALAM REGRESI. *Bina Ekonomi Majalah ilmiah Fakultas Ekonomi Unpar* (Volume 13, Nomor 1, Januari 2009).
- [23] Goldman, R. (2017). Foundations of Correlation and Regression. *Intermediate Biostatistics for Medical Researchers*.