

**TUGAS KELOMPOK**

**UJIAN AKHIR SEMSTER**

**“PREDIKSI HARGA PENJUALAN LAPTOP DENGAN MENGGUNAKAN METODE  
MULTIPLE LINIER REGRESSION DAN K-NEAREST NEIGHBOR”**

Disusun sebagai  
MATA KULIAH : MACHINE LEARNING



**Disusun Oleh**

Almira Rahma Sabita	2241727003
Dhika Ainul Lutfi	2241727007
Riris Silvia Zahri	2241727029

**KELAS TI-4J**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

**2022**

# **“PREDIKSI HARGA PENJUALAN LAPTOP DENGAN MENGGUNAKAN METODE LINIER REGRESSION DAN K-NEAREST NEIGHBOR”**

## **1. LATAR BELAKANG**

Laptop saat ini sudah menjadi sebuah kebutuhan, bukan lagi merupakan device yang exclusive. Kebutuhan setiap orang dalam menjalankan aktivitasnya baik kerja maupun belajar tidak lepas dari kebutuhan akan media komputer atau laptop. Pemasaran atau penjualan secara mudahnya adalah kegiatan memasarkan barang atau jasa umumnya kepada masyarakat dan khususnya kepada pembeli potensial. Saat ini, sebuah toko sebelum memasarkan produk harus merencanakan dan memprediksikan hasil dari penjualan atau dengan kata lain manajemen penjualan harus betul - betul bisa dianalisa dengan baik agar hasil penjualan yang didapat juga sesuai dengan harapan. Laptop juga memiliki kualitas, spesifikasi dan kemampuan yang berbeda beda yang bisa mempengaruhi harga jual kepada masyarakat. Penjual harus jeli dan hati hati dalam menentukan produk laptop merk dan spesifikasi yang bagaimana yang banyak diminati oleh konsumen.

Pada masa pandemi seperti ini, harga laptop ataupun elektronik mengalami peningkatan yang sangat signifikan. Untuk mendapatkan laba yang diinginkan, dibutuhkan manajemen harga jual laptop agar toko masih bisa mendapatkan laba hasil penjualan laptop dengan Metode Multiple Linier Regression Dan K-Nearest Neighbor untuk mengambil untuk prediksi atau meramal harga laptop. Dengan harapan dapat mempermudah atau mengetahui harga laptop pada bulan-bulan berikutnya.

## **2. METODE**

Dalam menentukan harga prediksi penjualan laptop menggunakan metode :

### **1. Metode Multiple Linier Regression**

Multiple Linear Regression adalah salah satu metode teknik statistik yang banyak digunakan memodelkan hubungan variabel Dependent dan Independent. Salah satu kegunaannya yaitu membuat model guna memprediksi nilai dari suatu variabel. Metode ini juga menampilkan hubungan sebab-akibat yang terjadi antar variable.

### **2. Metode K-Nearest Neighbor**

Metode algoritma K-Nearest Neighbor metode yang memiliki kemampuan untuk memprediksi nilai entitas yang tidak di ketahui, ini berguna dalam memprediksi harga laptop di masa mendatang berdasarkan data history sebelumnya.

### 3. PERHITUNGAN

#### A. Perhitungan Multiple Linier Regression

Multipel Linear regression merupakan salah satu metode statistik yang memberikan hasil output prediksi dengan melakukan pengembangan hubungan matematis antar variabel. Variabel independen merupakan variabel yang mempengaruhi sedangkan variabel dependen merupakan variabel yang dipengaruhi. **perhitungan multiple linear regression** yang digunakan maka dinyatakan pada.

$$Y = \alpha + b_1X_1 + b_2X_2 + \dots + b_n X$$

Y	=	variabel dependen (variabel tak terikat)
$X_1, X_2 \dots X$	=	Variabel independen (variabel terikat)
$\alpha$	=	Konstanta
$b_1, b_2 \dots b$	=	Koefisien regresi

Maka pernyataan diatas dapat dijadikan acuan untuk menghitung prediksi Prediksi Harga Penjualan Laptop dengan metode Multiple linear regression pada dapat dinyatakan dalam pernyataan

Y	=	Price_euros
X1	=	Inches
X2	=	Ram
X3	=	Weight

Kemudian mencari nilai  $X_1Y, X_2Y, X_3Y$ , Setelah mencari nilai tersebut ketemu maka mencari nilai  $X_1X_2, X_1X_3, X_2X_3, X_1^2, X_2^2, X_3^2$  dengan rumus.

$$A = \begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1X_2 & \Sigma X_1X_3 \\ \Sigma X_2 & \Sigma X_1X_2 & \Sigma X_2^2 & \Sigma X_2X_3 \\ \Sigma X_3 & \Sigma X_1X_3 & \Sigma X_2X_3 & \Sigma X_3^2 \end{bmatrix}$$

$$\hat{b}_0 = \frac{\text{Det } A_1}{\text{Det } A} \quad \hat{b}_2 = \frac{\text{Det } A_3}{\text{Det } A} \quad H = \begin{bmatrix} \Sigma Y \\ \Sigma X_1Y \\ \Sigma X_2Y \\ \Sigma X_3Y \end{bmatrix}$$

$$\hat{b}_1 = \frac{\text{Det } A_2}{\text{Det } A} \quad \hat{b}_3 = \frac{\text{Det } A_4}{\text{Det } A}$$

Setelah itu menjumlahkan rata rata di setiap variable yang sudah di cari sebelumnya dan variabel bebas.kemudian mencari nilai N yang berada pada matrix A pada kolom 1 baris 1.

Perhitungan Matrix A :

<b>Matriks A =</b>	1303	<b>19567,4</b>	<b>10922</b>	<b>2656,47</b>	<b>Matriks H =</b>	1464164,15
	19567,4	296496,1	166265	40915,52		22076158,47
	10922	166265	125212	23958,24		15711251,38
	2656,47	40915,5192	23958,24	5992,434		3112452,29
	<b>Det A =</b>	1,76E+13				
<b>Matriks A1</b>	1464164	19567,4	10922	2656,47	<b>Det A1 =</b>	2,14E+16
	22076158	296496,1	166265	40915,52		
	15711251	166265	125212	23958,24		
	3112452	40915,5192	23958,24	5992,434		
<b>Matriks A2</b>	1303	1464164,15	10922	2656,47	<b>Det A2 =</b>	-1,23E+15
	19567,4	22076158,47	166265	40915,52		
	10922	15711251,38	125212	23958,24		
	2656,47	3112452,29	23958,24	5992,434		
<b>Matriks A3</b>	1303	19567,4	1464164	2656,47	<b>Det A3 =</b>	1,84E+15
	19567,4	296496,1	22076158	40915,52		
	10922	166265	15711251	23958,24		
	2656,47	40915,5192	3112452	5992,434		
<b>Matriks A4</b>	1303	19567,4	10922	1464164	<b>Det A4 =</b>	6,59E+14
	19567,4	296496,1	166265	22076158		
	10922	166265	125212	15711251		
	2656,47	40915,5192	23958,24	3112452		

Hasil perhitungan dengan menghitung nilai Kofesien

$$\text{Koefisien } b_0 (\det A_1 / \det A) = 1220,334304$$

$$\text{Koefisien } b_1 (\det A_2 / \det A) = -70,1057959$$

$$\text{Koefisien } b_2 (\det A_3 / \det A) = 104,9384151$$

$$\text{Koefisien } b_3 (\det A_4 / \det A) = 37,53828669$$

$$Y \text{ Penduga(Predikted)} = 1220.334 - 70.1058X_1 + 104.9384X_2 + 37.53829X_3$$

Hasil Summary yang di dapatkan dalam perhitungan Multipel Linear regression

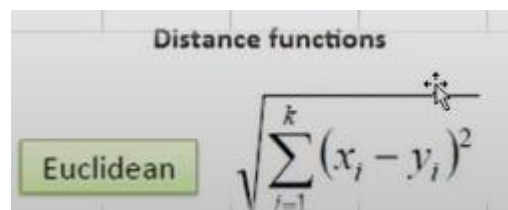
<i>Regression Statistics</i>	
Multiple R	0,751614589
R Square	0,564924491
Adjusted R Square	0,563919697
Standard Error	461,6002434
Observations	1303

## B. Perhitungan K-Nearest Neighbor

K-Nearest-Neighbor adalah algoritma yang bertujuan untuk mengklasifikasi objek data baru. Proses pengklasifikasikan objek data baru akan dilatih berdasarkan atribut dan data sample latih. Pemodelan klasifikasi objek baru hanya berdasarkan pada memori. Metode ini bekerja dengan mencari sejumlah k objek data (data latih) yang paling dekat dengan data uji yang diberikan, kemudian memilih kelas dengan jumlah voting terbanyak.

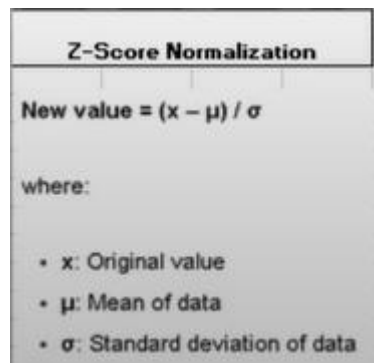
Langkah-langkah dari algoritma K-Nearest Neighbor sebagai berikut:

1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak Eucliden objek terhadap data training yang diberikan dengan rumus


$$\text{Euclidean} \quad \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

3. Selanjutnya mengurutkan hasil no 2 secara ascending (berurutan dari nilai tinggi ke rendah)

4. Melakukan normalisasi data dengan z score dengan rumus



**Z-Score Normalization**

New value =  $(x - \mu) / \sigma$

where:

- x: Original value
- $\mu$ : Mean of data
- $\sigma$ : Standard deviation of data

5. Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksi objek yang baru.

k pada algoritma k-nearest neighbor adalah banyaknya tetangga terdekat yang akan digunakan sebagai titik untuk melakukan klasifikasi pada data atau objek baru. Dalam menentukan jumlah nilai k, sebaiknya menggunakan angka ganjil. Untuk menghitung jarak antar objek data pada algoritma k-nearest neighbor dapat dilakukan dengan beberapa cara, salah satunya adalah dengan Ecludiean Distance.

Melakukan kalsifikasi kategori harga leptop dengan nilai rata rata yang di ambil sebagai patokan antara harga murah dan mahal.tabel di bawah ini contoh sebagai kasifikasi kategori data.

<b>laptop_ID</b>	<b>Inches</b>	<b>Ram</b>	<b>Weight</b>	<b>Price_euros</b>	<b>Kategori</b>
1	13,3	8	1,37	1339,69	Mahal
2	13,3	8	1,34	898,94	Murah
3	15,6	8	1,86	575	Murah
4	15,4	16	1,83	2537,45	Mahal
5	13,3	8	1,37	1803,6	Mahal

Data Testing

1321	13,5	6	2,2	1000	?
------	------	---	-----	------	---

Mentukan menentukan nilai normalisasi min dan max dengan rumus

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Dengan hasil nilai normalisasi min dan max dari data:

<b>Min(x)</b>	10,1	2	0,69	174
<b>Max(x)</b>	18,4	64	4,7	6099

#### 4. HASIL

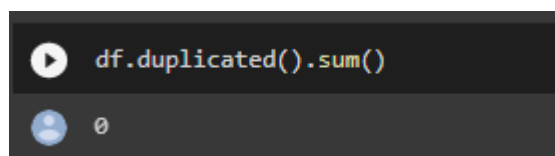
Hasil data set menggunakan Google Colab berbasis web base yang akan memudahkan dalam melakukan penelitian prediksi harga laptop. Google Colab adalah sebuah web aplikasi open-source adapun kegunaan lainnya dari Google Colab ini adalah untuk data cleaning, transformasi data, visualisasi data, machine learning.

	Laptop_ID	Company	Product	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price_euros	
	0	1	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	1339.69
	1	2	Apple	Macbook Air	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	898.94
	2	3	HP	250 G6	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	575.00
	3	4	Apple	MacBook Pro	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	2537.45
	4	5	Apple	MacBook Pro	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	1803.60

Gambar 1 Data Set Sebanyak 1304 Data.

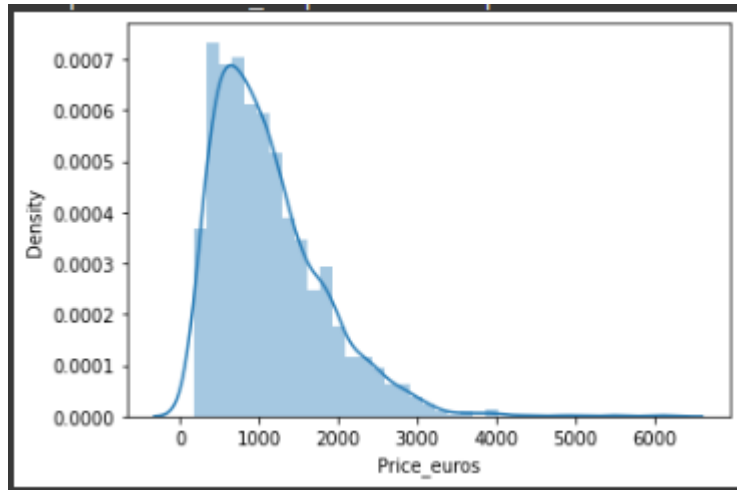
##### A. Hasil dengan Metode Multipel Linear regression

Dilakukan data cleaning pada tahap pre-processing. Kegunaannya adalah untuk memperbaiki atau menghapus data yang rusak maupun data yang tidak relevan. Selama masa pengumpulannya data yang terkumpul adalah 1304 data namun ditemukannya data duplicate, maka akan dilakukan data cleaning agar tidak memiliki data sama (data duplicate).



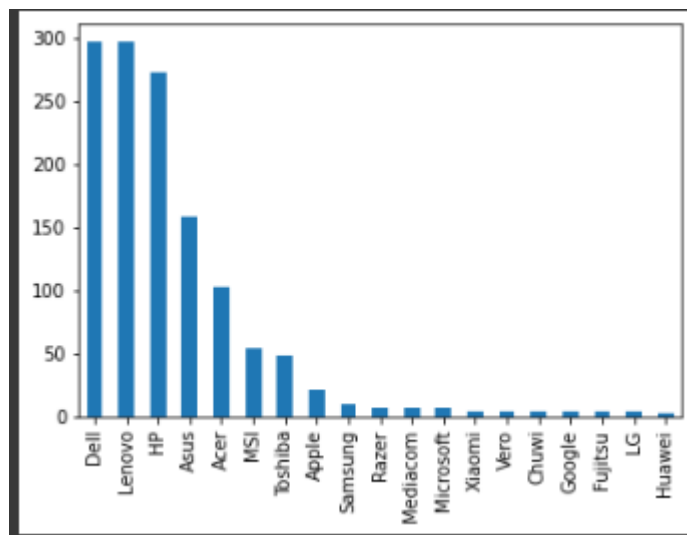
Gambar 2 Pengecekan Data Cleaning

Membuat visualisasi data merupakan hal yang agak tricky karena butuh keahlian dalam menyajikan data dengan baik. Data mentah atau raw data adalah data yang belum terorganisasi biasanya berupa tabel.



Gambar 3 Hasil Visualisasi Data

Berdasarkan hasil pada gambar 3 yang di dapat terlihat ada banyak laptop memiliki harga lebih murah dan beberapa laptop harganya sangat tinggi. Dilihat dari density maka merupakan hal yang jelas dimana pembeli laptop dengan harga rendah lebih banyak daripada dengan pembeli laptop harga tinggi.



Gambar 4 visualisasi data untuk melihat laptop yang paling diminati

Pada gambar 4Merk Dell & Lenovo memiliki jumlah pembeli yang tinggi. Lalu merk HP, Asus, Acer, MSI adalah jumlah laptop ke-2 tertinggi. Sedangkan merk yang lain memiliki pembeli dengan jumlah rendah.

Tahap untuk mengetahui besar pengaruh antara kedua variabel adalah dengan melakukan uji koefisien determinasi, sehingga dapat diketahui akurasi dari penerapan algoritma linear regression. Pada koefisien ini dengan nilai berkisar antara 0 sampai 1.



Jika nilai  $R^2 = 0$  dapat diartikan bahwa tidak adanya pengaruh atau hubungan antara kedua variabel (independent). dan dependen) dan jika  $R^2 = 1$  atau mendekati 1, maka memiliki hubungan yang semakin kuat antara variabel independen dengan variabel dependen

<b>R2 Score</b>
0.8244080785411312

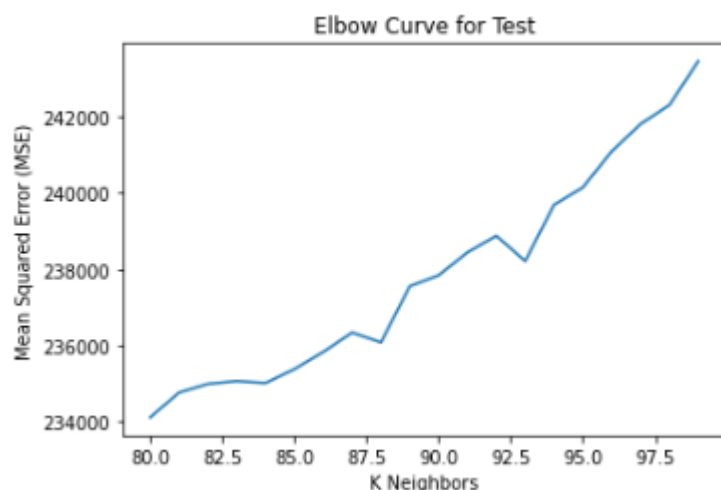
Dari hasil yang ditunjukkan di atas menunjukkan bahwa prediksi harga laptop metode linear regression, menghasilkan tingkat akurasi prediksi sebesar 0.8244080785411312 atau 82%.

<b>MAE</b>
0.20979730698399002

Selanjutnya adalah menguji nilai RMSE (Root Mean Square Error) untuk mengevaluasi hasil prediksi yang digunakan lalu mengukur nilai akurasi dari hasil penerapan prediksi pada model. Pada hasil di atas menunjukkan Root Mean Square Error prediksi sebesar 20979730698399002 atau Rp. 20979730698399002

## **B. Hasil dengan Metode K-Nearest Neighbor**

Pada Metode K-Nearest Neighbor, hasil visualisasi MSE menggunakan dataset sebanyak 1304 data dihasilkan MSE yaitu 234120.22857356857. Hasil visualisasi data terdekat dapat dilihat pada gambar di bawah ini.



Setelah melakukan pengujian menggunakan dataset sebanyak 1304 data dengan menggunakan Metode K-Nearest Neighbor dihasilkan akurasi tertinggi sebesar 53.154%. Hasil akurasi ini diperoleh dari setiap percobaan nilai k. Hasil akurasi dilihat pada tabel di bawah ini.

Akurasi
0.5315479153548515

## 5. KESIMPULAN

Dari hasil penerapan prediksi harga penjualan laptop dapat disimpulkan bahwa pengolahan awal data yang dilakukan pada data set 1304 data dalam tingkat akurasi prediksi harga penjualan laptop. Dengan algoritma multiple linear regression untuk memprediksi harga penjualan laptop dapat memberikan hasil keakuratan prediksi harga penjualan laptop dengan baik. Adapun hasil akurasi terbaik dengan menggunakan data untuk data testing menghasilkan tingkat akurasi prediksi sebesar 0.8244080785411312 atau 82%.. Namun, memiliki nilai galat cukup tinggi sebesar sebesar 20979730698399002 atau Rp. 20979730698399002.

Sedangkan dengan mengukana alogritma K-Nearest Neighbor diperoleh hasil akurasi akurasi tertinggi sebesar 53.154%, dengan nilai galat yaitu 234120.22857356857. Dari kedua hasil prediksi harga penjualan laptop dengan 2 metode multiple linear regression dan metode K-Nearest Neighbor maka untuk data prediksi lebih bagus dan lebih tepat menggunakan metode Metode Multiple Linier Regression.