

## MACHINE LEARNING UNTUK MODEL PREDIKSI HARGA SEMBAKO DENGAN METODE REGRESI LINIER BERGANDA

<sup>1)</sup>Kandari Puteri, <sup>2)</sup>Astried Silvanie

<sup>1)2)</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer

<sup>1)2)</sup>Institut Bisnis dan Informatika (IBI) Kosgoro 1957

Email: <sup>1)</sup>kandariputeri@gmail.com, <sup>2)</sup>astried@ibi-k57.ac.id

### ABSTRAK

Harga sembilan bahan pokok (sembako) setiap waktu dapat naik dan turun (fluktuatif), serta kebutuhan akan informasi harga sembako harian. Oleh karena itu diperlukannya peramalan harga sembako harian untuk beberapa waktu ke depan. Salah satu metode yang dapat digunakan untuk memprediksi harga yang bernilai numerik kontinu yaitu dengan menggunakan metode regresi. Penelitian ini bertujuan untuk mengimplementasikan metode regresi linier berganda dalam memprediksi harga sembako harian, menggunakan sampel data sembako di DKI Jakarta. Dengan himpunan data selama 4 tahun terakhir, yaitu dari tanggal 1 Januari 2016 sampai dengan tanggal 31 Desember 2019, yang diperoleh dari portal data Pemerintah Provinsi DKI Jakarta dialamat *website* [www.data.jakarta.go.id](http://www.data.jakarta.go.id). Terdapat 4 variabel di dalam himpunan data tersebut yaitu, variabel tanggal, komoditas, pasar dan harga. Variabel tanggal, komoditas dan pasar merupakan variabel bebas (*predictor*), sedangkan variabel harga merupakan variabel terikat (*response*). Persentase sumbangan pengaruh variabel bebas terhadap variabel terikat sebesar 84,2%, sedangkan sisanya sebesar 15,8% dipengaruhi oleh variabel yang tidak dimasukkan ke dalam penelitian ini.

**Kata kunci:** **Prediksi, Sembilan Bahan Pokok, Regresi Linier Berganda, Machine Learning, Data Mining**

### I. PENDAHULUAN

Sembilan bahan pokok (sembako) merupakan kebutuhan yang wajib dijual bebas dipasar untuk memenuhi kebutuhan konsumsi sehari-hari. Harga sembako setiap waktu dapat naik dan turun (fluktuatif) disebabkan dari faktor eksternal yang terus berubah, serta kebutuhan akan informasi harga sembako harian. Oleh karena itu diperlukannya peramalan harga sembako harian untuk beberapa waktu yang akan datang.

Langkah yang dapat dilakukan untuk mengatasi permasalahan tersebut yaitu dengan membuat sistem prediksi harga sembako harian. Salah satu metode yang dapat digunakan untuk memprediksi harga yang bernilai numerik kontinu yaitu dengan menggunakan metode regresi. Regresi bertujuan menemukan suatu fungsi yang memodelkan data. Proses prediksi dalam *machine learning* yaitu dengan memahami sifat atau ciri dari objek yang tidak dikenal dengan cara mengidentifikasi pola (*pattern*) dalam himpunan data (*dataset*).

Ciri khas dari *machine learning* yaitu adanya pelatihan, pembelajaran atau *training*. Oleh karena itu *machine learning* membutuhkan data untuk dipelajari sebagai data latih (*training set*), dalam hal ini yaitu variabel tanggal, komoditas, pasar dan harga. Model yang akan dihasilkan dari proses pelatihan akan dijadikan acuan dalam memprediksi harga sembako harian. Serta pendekatan *data mining*, dimana penelitian ini dilakukan penambangan dalam himpunan data yang terdiri dari 90.945 baris data yang nantinya akan menghasilkan informasi baru yang bermanfaat, yaitu prediksi harga sembako harian. Menggunakan bahasa pemrograman *python* sebagai *backend*, dan *web framework django*.

Adapun penelitian sebelumnya yang telah dilakukan oleh Nafi'iyah (2015) menghasilkan persamaan yang diperoleh dari data mobil Innova tanpa normalisasi yaitu  $y = -0,00000000038632 + 0,0000074163x_1 + 0,00000019242x_2 + 0,4393x_3$ . Dan persamaan yang

diperoleh dari data mobil Innova dengan normalisasi terlebih dahulu yaitu  $y = -2,6293 + 0,2812x_1 + 0,8105x_2 + 0,2013x_3$ . Hasil koefisien dari korelasi dari cara pertama, yaitu  $R^2 = -0,68$ . Dan hasil koefisien korelasi dari cara kedua, yaitu  $R^2 = 0,93373$ .

## II. METODE PENELITIAN

### 2.1 Prediksi

Prediksi (*prediction*) adalah memperkirakan nilai-nilai data bertipe apa saja dan kapan saja (masa lalu, sekarang, dan masa depan). Terdapat satu istilah yang mirip dengan prediksi, yaitu peramalan (*forecasting*) adalah memperkirakan nilai-nilai data *time series* dimasa depan (Suyanto, 2018).

### 2.2 Sembilan Bahan Pokok

Sembilan bahan pokok (sembako) merupakan kebutuhan wajib yang dijual bebas dipasar untuk memenuhi kebutuhan sehari-hari. Sembako adalah sembilan jenis kebutuhan pokok masyarakat menurut Peraturan Menteri Perdagangan Republik Indonesia Nomor 07 Tahun 2020 Tentang Harga Acuan Pembelian di Tingkat Petani dan Harga Acuan Penjualan di Tingkat Konsumen (disingkat: "Permendag No. 07 Th. 2020"). Kesembilan bahan pokok itu yaitu Beras, Gula Pasir, Minyak Goreng dan Mentega, Daging Sapi dan Daging Ayam, Telur Ayam, Susu, Jagung, Gas dan Minyak Tanah, serta Garam Beryodium (Permendag, No. 07 Th. 2020).

### 2.3 Regresi Linier Berganda

Model regresi linier berganda (*multiple linear regression*) mengasumsikan sebuah hubungan linier (dalam parameter) antara sebuah hubungan linier (dalam parameter) antara sebuah variabel respon  $y$  (*dependent variable*) dan sebuah himpunan dari variabel prediktor (*explanatory/independent variable*)  $x = (x_1, x_2, \dots, x_k)$ . Model regresi linier berganda dengan  $k$  variabel prediktor  $x_1, x_2, \dots, x_k$  dan satu variabel respon  $y$  (Schmidheiny, 2019), persamaan regresi linier berganda dapat ditulis sebagai berikut:  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon$  ..... (1)

#### 2.3.1 Uji Asumsi Klasik

Uji asumsi klasik adalah persyaratan statistik yang harus dipenuhi pada analisis regresi linier berganda. Koefisien-koefisien regresi linier sebenarnya adalah nilai duga dari parameter model regresi. Parameter merupakan keadaan sesungguhnya untuk kasus yang diamati. Untuk mengetahui koefisien regresi yang didapatkan telah sah (benar; dapat diterima), maka diperlukan pengujian terhadap kemungkinan adanya pelanggaran asumsi klasik tersebut (Gujarati, 2003; Kurniawan, 2008). Asumsi klasik regresi linier berganda adalah sebagai berikut:

- 1) Model dispesifikasi dengan benar
- 2) Model regresinya adalah linier dalam variabel dan parameter
- 3) *Error* menyebar dengan normal
- 4) Ragam dari *error* bersifat homogen (homoskedastisitas)
- 5) Tidak terjadinya multikolinieritas antara variabel bebas  $X$
- 6) *Error* tidak mengalami autokorelasi

#### 2.3.2 Uji Hipotesis

Regresi linier berganda merupakan salah satu pengujian hipotesis untuk mengetahui pengaruh antara variabel bebas (*independent*) terhadap variabel terikatnya (*dependent*). Sehingga yang dicari adalah pengaruh variabel bebas yaitu tanggal ( $X_1$ ), komoditas ( $X_2$ ) dan pasar ( $X_3$ ) terhadap variabel terikat yaitu harga ( $Y$ ). Pengujian yang dapat dilakukan adalah pengujian parsial, dalam pengujian

ini ingin diketahui apakah jika secara terpisah, suatu variabel  $X$  masih memberikan kontribusi secara signifikan terhadap variabel terikat  $Y$  (Kurniawan, 2008).

### 2.3.3 Uji Kelayakan Model

Pada tahap ini dilakukan pengujian meliputi uji simultan model regresi dan koefisien determinasi.

1) Uji simultan model regresi

Uji simultan (keseluruhan; bersama-sama) pada konsep regresi linier adalah pengujian mengenai apakah model regresi yang didapatkan benar-benar dapat diterima. Uji simultan bertujuan untuk menguji apakah antara variabel-variabel bebas  $X$  dan terikat  $Y$ , atau setidaknya tidaknya antara satu variabel  $X$  dengan variabel terikat  $Y$ , benar-benar terdapat hubungan linier (*linear relation*) (Kurniawan, 2008).

2) Koefisien determinasi

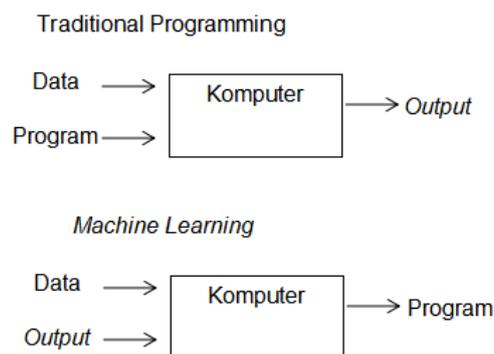
Koefisien determinasi adalah besarnya keragaman (informasi) di dalam variabel  $Y$  yang dapat diberikan oleh model regresi yang didapatkan. Semakin besar nilai  $R^2$ , semakin baik model regresi yang diperoleh (Kurniawan, 2008). Berikut ini merupakan persamaan dari  $R^2$ , yaitu:

$$R^2 = \frac{(\sum y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i^2)} \dots\dots\dots (2)$$

## 2.4 Machine Learning

Pembelajaran Mesin (*Machine Learning*) merupakan salah satu cabang dari ilmu Kecerdasan Buatan (*Artificial Intelligence*), khususnya yang mempelajari tentang bagaimana komputer mampu belajar dari data untuk meningkatkan kecerdasannya (Wahyono, 2018).

Menurut Mitchell (1997) komputer yang memiliki kemampuan melakukan belajar dari pengalaman terhadap tugas-tugasnya dan mengalami peningkatan kinerja. Sedangkan, menurut Arthur (1959) kemampuan komputer untuk melakukan pembelajaran tanpa harus menjelaskan atau terprogram secara eksplisit kepada komputer.



**Gambar 1. Perbedaan Pemrograman Tradisional dan Pembelajaran Mesin (Brownlee, 2015; Chollet, 2018).**

Pada pemrograman tradisional, data dan program dijalankan pada komputer untuk menghasilkan keluaran, sedangkan pada pemrograman menggunakan pembelajaran mesin, data dan keluaran dijalankan pada komputer untuk membuat program, dan kemudian program tersebut bisa digunakan dalam pemrograman tradisional (Brownlee, 2015; Chollet, 2018).

Harrington (2012) memaparkan beberapa proses untuk membangun sebuah sistem *machine learning*, yaitu Pengumpulan Data (*Collect Data*), Mempersiapkan Data Masukkan (*Prepare Input Data*), Menganalisis Data Masukkan (*Analyse Input Data*), Mengikutsertakan Keterlibatan Manusia

(*Human Involvement*), Melatih Algoritma (*Train Algorithm*), Menguji Algoritma (*Test Algorithm*), dan yang terakhir Menggunakan Model (*Use It*).

## 2.5 Data Mining

Penambangan data (*data mining*) juga dikenal sebagai menemukan pengetahuan dari data (*knowledge discovery from data*). *Data mining* dapat memiliki dengan tidak sah dinamai “*knowledge mining from data*” atau “*knowledge mining*” yaitu pada penambangan dari data dengan jumlah sangat besar. *Data mining* dapat dilihat sebagai sebuah hasil dari evolusi natural dari teknologi informasi. Basisdata dan industri manajemen data, berkembang secara lambat di pengembangan masing-masing fungsionalitas kritis: *data collection and database creation*, *data management* (termasuk penyimpanan data dan pengambilan kembali dan pemrosesan transaksi basisdata), dan kemajuan analisis data (termasuk *data warehousing* dan *data mining*) (Han, 2012).

## III. HASIL DAN PEMBAHASAN

### 3.1 Himpunan Data

Penelitian ini menggunakan himpunan data selama 4 tahun terakhir, yaitu dari tanggal 1 Januari 2016 sampai dengan tanggal 31 Desember 2019, yang diperoleh dari portal data Pemerintah Provinsi DKI Jakarta di alamat *website* [www.data.jakarta.go.id](http://www.data.jakarta.go.id). Variabel “tanggal, komoditas dan pasar” merupakan variabel bebas (*predictor*), sedangkan variabel “harga” merupakan variabel terikat (*response*). Variabel “tanggal” bernilai *datetime*, variabel “komoditas dan pasar” bernilai nominal, dan variabel “harga” bernilai skala interval. Menggunakan total data penelitian sebanyak 90.945 baris data. Berikut ini contoh data sembako dari Pemprov DKI Jakarta yaitu:

**Tabel 1. Contoh Data Sembako**

| Tanggal  | Komoditas         | Pasar             | Harga |
|----------|-------------------|-------------------|-------|
| 1/1/2016 | Cabai Rawit Merah | Pasar Jatinegara  | 40000 |
| 1/1/2016 | Bawang Merah      | Pasar Jatinegara  | 35000 |
| 1/1/2016 | Bawang Putih      | Pasar Kramat Jati | 25000 |
| 1/1/2016 | Gas Elpiji 3kg    | Pasar Minggu      | 20000 |

### 3.2 Pra-pemrosesan Data

Sebelum melakukan proses estimasi model dilakukan pra-pemrosesan data yang bertujuan untuk mengurangi komputasi yang dilakukan mesin, mempercepat waktu dan meningkatkan akurasi dari proses estimasi model. Dalam proses pra-pemrosesan data ada beberapa tahap yang dilakukan antara lain:

#### 3.2.1 Pembersihan Data

##### 1) Mengisi Nilai Variabel yang Kosong (*Missing Value*)

Karena jumlah nilai variabel yang kosong pada penelitian ini hanya sedikit, maka pengisian dilakukan secara manual. Ditunjukkan pada gambar 2 merupakan contoh nilai variabel yang kosong.

| tanggal   | komoditas      | pasar            | harga |
|-----------|----------------|------------------|-------|
| 1/25/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1/26/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1/27/2016 | Gas Elpiji 3kg | Pasar Jatinegara |       |
| 1/28/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1/29/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |

**Gambar 2. Contoh Nilai Variabel Kosong**

## 2) Memperbaiki Data yang Tidak Konsisten

Dalam himpunan data penelitian ini juga terdapat data yang tidak konsisten, seperti pada kolom harga pada gambar 3 di bawah ini:

| tanggal  | komoditas      | pasar            | harga |
|----------|----------------|------------------|-------|
| 3/1/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 19000 |
| 3/2/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 19000 |
| 3/3/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 1900  |
| 3/4/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 19000 |
| 3/5/2016 | Gas Elpiji 3kg | Pasar Jatinegara | 19000 |

**Gambar 3. Contoh Data Tidak Konsisten**

## 3) Merubah Tipe Data Pada Variabel

Pada analisis regresi linier diharuskan data masukan bernilai numerik agar dapat melalui proses perhitungan, untuk itu perlu dilakukan perubahan tipe data pada variabel “tanggal” yang bertipe *datetime* menjadi *timestamp* dan variabel “komoditas” yang bertipe nominal menjadi interval. Pada gambar 4 variabel “tanggal” setelah diubah menjadi tipe data *timestamp*, yaitu:

| tanggal    | komoditas      | pasar            | harga |
|------------|----------------|------------------|-------|
| 1451581200 | Gas Elpiji 3kg | Pasar Jatinegara | 17500 |
| 1451667600 | Gas Elpiji 3kg | Pasar Jatinegara | 17500 |
| 1451754000 | Gas Elpiji 3kg | Pasar Jatinegara | 17500 |
| 1451840400 | Gas Elpiji 3kg | Pasar Jatinegara | 17500 |
| 1451926800 | Gas Elpiji 3kg | Pasar Jatinegara | 17500 |
| 1452013200 | Gas Elpiji 3kg | Pasar Jatinegara | 17500 |
| 1452099600 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1452186000 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1452272400 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1452358800 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1452445200 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1452531600 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |
| 1452618000 | Gas Elpiji 3kg | Pasar Jatinegara | 18000 |

**Gambar 4. Variabel “tanggal” Bernilai *Timestamp***

Pada variabel “komoditas” yang sebelumnya bernilai nominal diubah menjadi tipe data interval, dapat dilihat pada tabel 2 di bawah ini:

**Tabel 2. Skala Interval Variabel Komoditas**

| Komoditas                  | Interval |
|----------------------------|----------|
| Daging Sapi Murni (Semur)  | 23       |
| Daging Kambing             | 22       |
| Cabai Rawit Merah          | 21       |
| Cabai Rawit Hijau          | 20       |
| Cabai Merah Besar          | 19       |
| Cabai Merah Keriting       | 18       |
| Susu Bubuk Dancow 400kg    | 17       |
| Bawang Putih               | 16       |
| Bawang Merah               | 15       |
| Ikan Bandeng (Sedang)      | 14       |
| Ayam                       | 13       |
| Telur Ayam                 | 12       |
| Gas Elpiji 3kg             | 11       |
| Kentang (Sedang)           | 10       |
| Minyak Goreng              | 9        |
| Beras IR. II (IR 64) Ramos | 8        |

|                       |   |
|-----------------------|---|
| Jagung Pipilan        | 7 |
| Gula Pasir            | 6 |
| Kacang Kedelai        | 5 |
| Margarin Blueband Cup | 4 |
| Tepung Terigu         | 3 |
| Ketela Pohon          | 2 |
| Garam Dapur           | 1 |

Pada variabel “pasar” yang sebelumnya bernilai nominal diubah menjadi tipe data interval, dapat dilihat pada tabel 3 di bawah ini:

**Tabel 3. Skala Interval Variabel Pasar**

| Pasar             | Interval |
|-------------------|----------|
| Pasar Jatinegara  | 1        |
| Pasar Kramat Jati | 2        |
| Pasar Minggu      | 3        |

Setelah melalui proses pengisian nilai variabel yang kosong (*missing values*), memperbaiki data yang tidak konsisten, serta mengubah tipe data pada variabel “tanggal, komoditas dan pasar”. Berikut ini merupakan data sembako setelah melalui proses pembersihan data, yaitu:

| tanggal    | komoditas | pasar | harga  |
|------------|-----------|-------|--------|
| 1451581200 | 23        | 1     | 120000 |
| 1451667600 | 23        | 1     | 120000 |
| 1451754000 | 23        | 1     | 120000 |
| 1451840400 | 23        | 1     | 120000 |
| 1451926800 | 23        | 1     | 120000 |
| 1452013200 | 23        | 1     | 120000 |
| 1452099600 | 23        | 1     | 120000 |
| 1452186000 | 23        | 1     | 120000 |
| 1452272400 | 23        | 1     | 120000 |
| 1452358800 | 23        | 1     | 120000 |

**Gambar 5. Data Sembako Setelah Proses Pembersihan Data**

### 3.2.2 Transformasi Data

Transformasi data dilakukan dengan tujuan utama untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga dapat memenuhi asumsi-asumsi yang mendasari analisis ragam. Serta, data ditransformasikan agar menghasilkan proses perhitungan yang lebih efisien. Dalam penelitian ini menggunakan *power transform*.

#### 1) *Power Transform*

Sebuah *power transform* menghapus sebuah pergeseran dari distribusi data untuk membuat distribusi lebih normal (*Gasussian*). Pada *power transform* ini dapat diimplementasikan dengan transformasi *Box-Cox* dalam *Python* menggunakan fungsi `boxcox()` dari pustaka *Scipy*. Berikut pada gambar 6 hasil dari transformasi *boxcox*.

```
Out [20]:
```

|   | tanggal   | komoditas | pasar | harga     |
|---|-----------|-----------|-------|-----------|
| 0 | 21.095919 | 3.135494  | 0.0   | 11.695247 |
| 1 | 21.095979 | 3.135494  | 0.0   | 11.695247 |
| 2 | 21.096038 | 3.135494  | 0.0   | 11.695247 |
| 3 | 21.096098 | 3.135494  | 0.0   | 11.695247 |
| 4 | 21.096157 | 3.135494  | 0.0   | 11.695247 |

**Gambar 6. Hasil Transformasi *Boxcox***

### 3.3 Estimasi Model

#### 3.3.1 Pembelajaran Model Regresi Linier

Selanjutnya, proses pembelajaran model pada himpunan data. Estimator yang digunakan adalah regresi linier (*linear regression*). Berikut pada gambar 7 *output* pembelajaran model regresi linier.

```
Out[9]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

**Gambar 7. Output Pembelajaran Model Regresi Linier**

Pada gambar 7 menunjukkan pesan keluaran setelah proses pembelajaran model regresi linier pada himpunan data ( $X$  dan  $Y$ ).

Setelah proses pembelajaran model regresi linier, didapatkan koefisien-koefisien regresi linier sebagai berikut:

```
Intersep : -1.5609145371353428
Slope
Tanggal : 0.42831111858781074
Komoditas : 1.0921548656902518
Pasar : 0.015082588337265642
```

**Gambar 8. Intersep dan Slope**

Berdasarkan hasil dari pembelajaran regresi linier berganda menghasilkan model regresi sebagai berikut:  $Y = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ..... (3)

Menjadi:  $Y = -1.5609 + 0.4283(X_1) + 1.0922(X_2) + 0.0151(X_3) + \varepsilon$  ..... (4)

Dimana:

- $Y$  = Harga (variabel terikat)
- $a$  = Konstanta/Intersep
- $\beta_1$  = Slope tanggal
- $\beta_2$  = Slope komoditas
- $\beta_3$  = Slope pasar
- $X_1, X_2, X_3$  = Variabel tanggal, komoditas, dan pasar
- $\varepsilon$  = *Error*

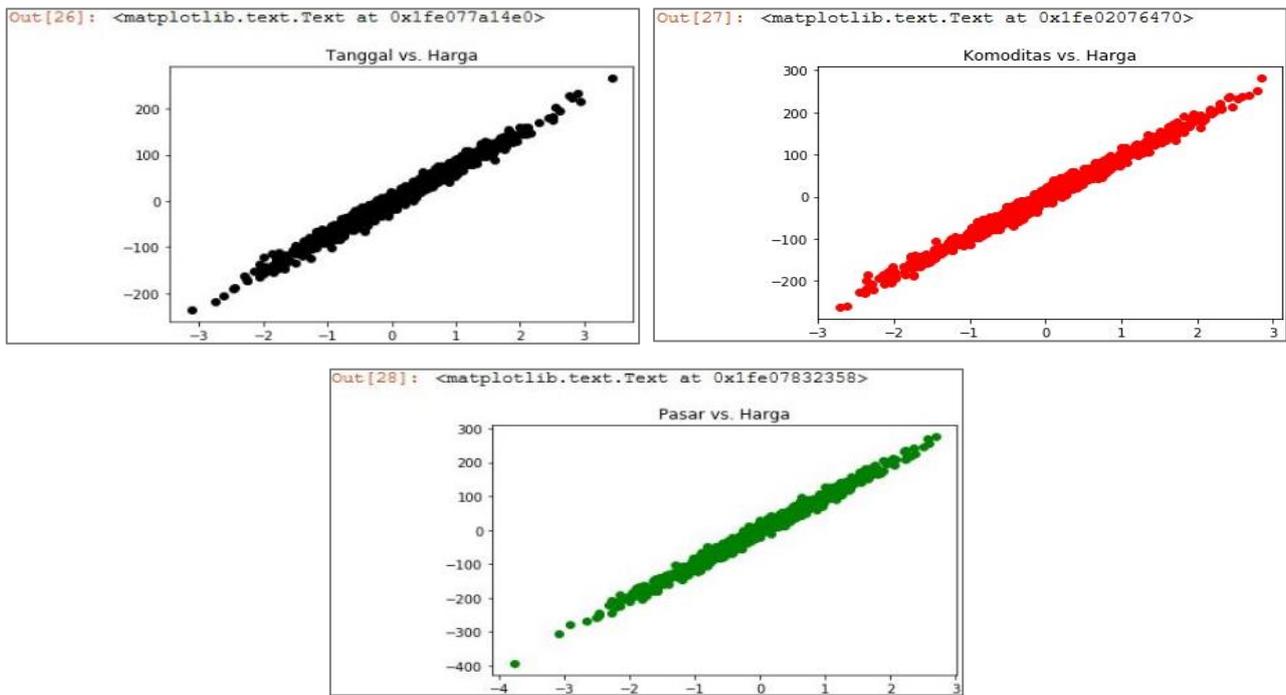
Dengan demikian pembelajaran mesin dengan model regresi linier sudah siap dijalankan.

### 3.4 Uji Asumsi Klasik

#### 3.4.1 Linieritas

Suatu model regresi harus memiliki hubungan yang linier dalam parameter dan variabel. Model regresi linier dalam parameter, jika  $\beta_i$  memperlihatkan dengan hanya satu dan tidak digandakan atau dibagi oleh parameter lainnya, dalam interpretasi ini  $E(Y | X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$  adalah sebuah model regresi linier dalam parameter.

Sedangkan, model regresi linier dalam variabel yaitu dapat ditunjukkan dengan hubungan linier antara variabel bebas dan variabel terikat. Berikut pada gambar 9 hasil uji linieritas.

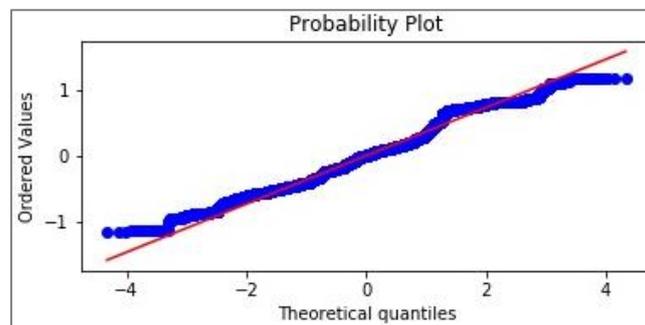


**Gambar 9. Hasil Uji Linieritas**

Berdasarkan hasil pengujian linieritas, disimpulkan bahwa antara variabel bebas dan terikat memiliki hubungan yang linier.

### 3.4.2 Normalitas Residual

Uji asumsi klasik selanjutnya yaitu *error*/residual menyebar dengan normal. Pengujian kenormalan residual dapat dilakukan dengan QQ-Plot, berikut pada gambar 10 grafik normalitas residual.



**Gambar 10. Grafik Normalitas Residual**

Pada grafik normalitas residual di atas terlihat bahwa titik-titik data tersebut tersebar di sekitar garis lurus, jika dapat didekati atau digambarkan dengan garis lurus maka data tersebut menyebar dengan normal.

### 3.4.3 Homoskedastisitas

Ragam dari *error* bersifat homogen (homoskedastic), yaitu *Error* memiliki nilai ragam yang sama antara *error* ke-*i* dan *error* ke-*j*. Statistik uji yang digunakan adalah uji Goldfeld-Quandt. Pada gambar 4.37 hasil pengujian homoskedastisitas. Dengan menggunakan statistik uji Goldfeld-Quandt, hasil yang diperoleh adalah sebagai berikut:

```
Out[17]: [('F statistic', 0.4358602705939647), ('p-value', 0.9999999999999999)]
```

**Gambar 11. Hasil Uji Homoskedastisitas**

Statistik uji Goldfeld-Quandt menghasilkan nilai p-value yang lebih besar dari 0.05 yang mengindikasikan penerimaan  $H_0$ . Maka, dapat disimpulkan bahwa tidak terjadi heteroskedastisitas pada ragam *error*.

### 3.4.4 Non-Multikolinieritas

Multikolinieritas berarti bahwa terjadi korelasi linier yang erat antar variabel bebas. Uji statistik yang digunakan adalah dengan *Variance Inflation Factor (VIF)*. Nilai VIF yang lebih besar dari 10 mengindikasikan adanya multikolinieritas yang serius. Pada gambar 4.38 hasil pengujian non-multikolinieritas. Hasil uji asumsi non-multikolinieritas dengan menggunakan statistik VIF adalah sebagai berikut:

```
Out[16]:
```

|   | Variabel  | VIF      |
|---|-----------|----------|
| 0 | tanggal   | 0.000474 |
| 1 | komoditas | 0.995960 |
| 2 | pasar     | 0.996653 |

**Gambar 12. Hasil Uji Non-Multikolinieritas**

Dari hasil uji non-multikolinieritas di atas nilai VIF dari ketiga variabel bebas lebih kecil dari 10, hal ini menunjukkan bahwa tidak terjadi multikolinieritas di antara variabel bebas.

### 3.4.4 Non-Autokorelasi

Autokorelasi adalah korelasi yang terjadi antara observasi dalam satu variabel. Untuk menguji autokorelasi pada data residual dapat menggunakan fungsi *Pandas.Series.autocorr()* yang mengembalikan nilai koefisien korelasi *pearson*.

```
Out[29]: 0.06291198932822399
```

**Gambar 13. Hasil uji non-autokorelasi**

Hasil dari uji autokorelasi pada gambar 13 menghasilkan nilai 0, ini mengindikasikan tidak ada autokorelasi.

## 3.5 Uji Hipotesis

### 3.5.1 Uji Parsial (Uji Statistik-T)

Uji - T digunakan untuk mengetahui apakah dalam model regresi variabel bebas ( $X_1$ ,  $X_2$ , dan  $X_3$ ) yaitu tanggal, komoditas, dan pasar secara parsial berpengaruh secara parsial berpengaruh secara signifikan terhadap variabel terikat ( $Y$ ) yaitu harga.

```
Tanggal
T-Statistic : 268527.92504054337
P-Value : 0.0

Komoditas
T-Statistic : 899.9638913681611
P-Value : 0.0

Pasar
T-Statistic : 397.07197291697855
P-Value : 0.0
```

**Gambar 14. Hasil T-Hitung Variabel Bebas**

Dalam uji-t, akan dirumuskan hipotesa awal dan hipotesa alternatifnya. Dimana hipotesa tersebut adalah :

$H_0$  = Secara parsial tidak terdapat pengaruh secara signifikan antara tanggal, komoditas dan pasar terhadap harga.

$H_1$  = Secara parsial terdapat pengaruh secara signifikan antara tanggal, komoditas dan pasar terhadap harga.

Berdasarkan pada gambar 14 Hasil t-hitung variabel bebas, pada kolom  $t$  dapat diketahui bahwa nilai  $T_{hitung}$  untuk tanggal adalah 268527.925, komoditas adalah 899.964, dan pasar adalah 397.072. Serta, nilai  $T_{tabel}$  dengan derajat kebebasan ( $df$ ), untuk  $df$  sebesar 90941 dengan nilai  $alpha$  (tingkat kepercayaan) 5%, sehingga nilai  $T_{tabel}$  adalah 1.959990071. Sehingga  $T_{hitung} > T_{tabel}$  untuk :

- 1)  $X_1$  yaitu sebesar  $268527.925 > 1.959990071$ , sehingga  $H_0$  ditolak. Dengan demikian terbukti bahwa tanggal ( $X_1$ ) secara parsial berpengaruh signifikan terhadap harga ( $Y$ ).
- 2)  $X_2$  yaitu sebesar  $899.964 > 1.959990071$ , sehingga  $H_0$  ditolak. Dengan demikian terbukti bahwa komoditas ( $X_2$ ) secara parsial berpengaruh signifikan terhadap harga ( $Y$ ).
- 3)  $X_3$  yaitu sebesar  $397.072 > 1.959990071$ , sehingga  $H_0$  ditolak. Dengan demikian terbukti bahwa pasar ( $X_3$ ) secara parsial berpengaruh signifikan terhadap harga ( $Y$ ).

### 3.6 Uji Kelayakan Model

#### 3.6.1 Uji Simultan Model Regresi (Uji Statistik-F)

Uji - F digunakan untuk mengetahui apakah variabel bebas ( $X_1$ ,  $X_2$ , dan  $X_3$ ) yaitu tanggal, komoditas, dan pasar mempengaruhi secara signifikan terhadap variabel terikat ( $Y$ ) yaitu harga. Berikut pada gambar 15 Hasil pengolahan data untuk uji simultan model regresi.

Out [18] :

|           | sum_sq       | df      | F             | PR(>F)       |
|-----------|--------------|---------|---------------|--------------|
| tanggal   | 9.399089     | 1.0     | 68.872813     | 1.064404e-16 |
| komoditas | 66304.232149 | 1.0     | 485851.218504 | 0.000000e+00 |
| pasar     | 4.256798     | 1.0     | 31.192136     | 2.343764e-08 |
| Residual  | 12410.740050 | 90941.0 | NaN           | NaN          |

Gambar 15. Hasil Uji Simultan Model Regresi

Dalam uji-f, akan dirumuskan hipotesa awal dan hipotesa alternatif. Dimana hipotesa tersebut adalah:

$H_0$  = Tidak terdapat pengaruh secara signifikan antara tanggal, komoditas dan pasar terhadap harga.

$H_1$  = Terdapat pengaruh secara signifikan antara tanggal, komoditas dan pasar terhadap harga.

Berdasarkan gambar 15, dapat diketahui bahwa nilai  $F_{hitung}$  untuk tanggal adalah 68.872813, komoditas adalah 485851.218504, dan pasar adalah 31.192136. Serta, nilai  $F_{tabel}$  dengan derajat kebebasan ( $df$ ), untuk  $df1$  sebesar 3 dan  $df2$  sebesar 90942 dengan nilai  $alpha$  (tingkat kepercayaan) 5%, sehingga nilai  $F_{tabel}$  adalah 2.605006903. Sehingga  $F_{hitung} > F_{tabel}$ , maka  $H_0$  ditolak. Dengan demikian terbukti bahwa tanggal ( $X_1$ ), komoditas ( $X_2$ ), dan pasar ( $X_3$ ) secara bersama-sama mempunyai pengaruh yang signifikan terhadap harga ( $Y$ ). Maka model regresi yang diperoleh dapat digunakan untuk memprediksi harga sembako.

### 3.6.2 Koefisien Determinasi

Koefisien determinasi adalah besarnya keragaman (informasi) di dalam variabel Y yang dapat diberikan oleh model regresi yang didapatkan.

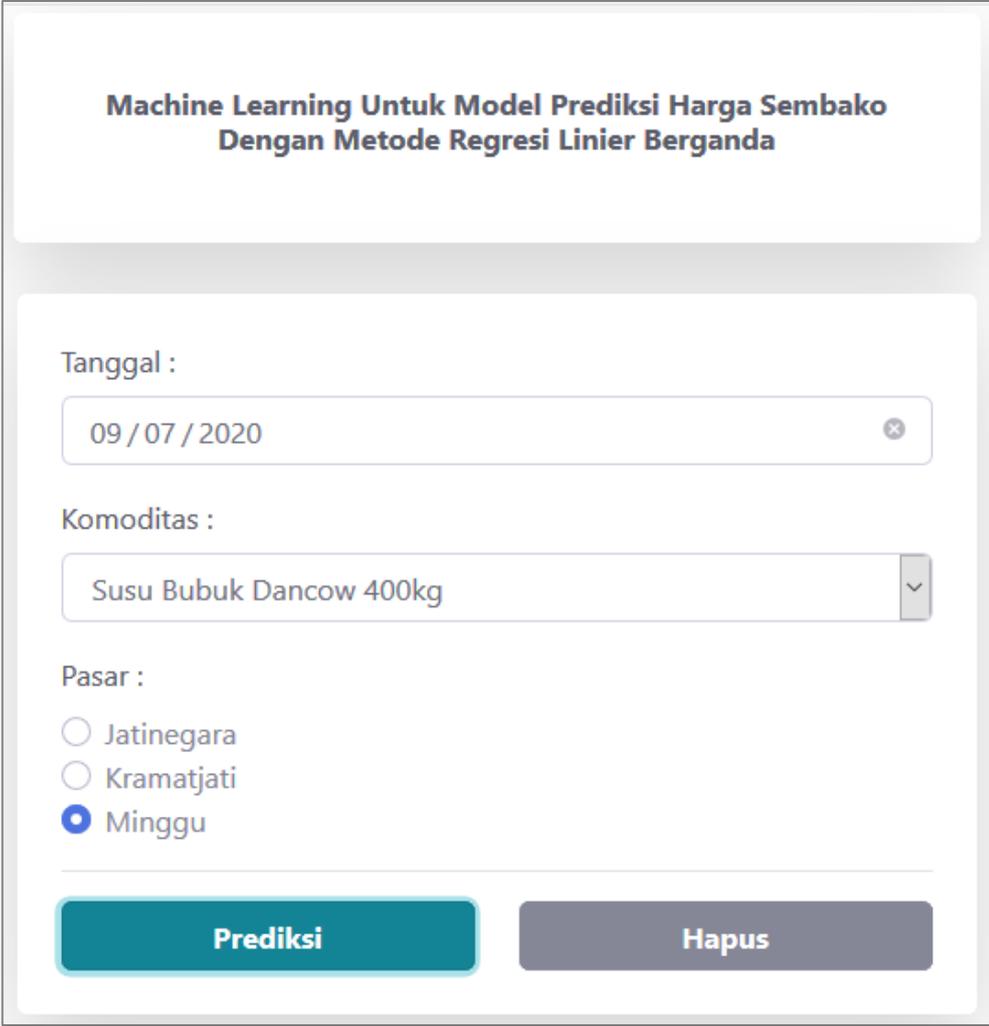
R-Squared : 0.8423484242569402

**Gambar 16. Nilai *r-squared***

Berdasarkan pada gambar 16, dapat diketahui bahwa nilai  $R^2$  (*R-squared*) sebesar 0.842 atau 84,2%. Hal ini menunjukkan bahwa persentase sumbangan pengaruh variabel bebas ( $X_1$ ,  $X_2$ , dan  $X_3$ ) yaitu tanggal, komoditas dan pasar terhadap variabel terikat ( $Y$ ) yaitu harga sebesar 84,2%. Sedangkan sisanya sebesar 15,8% dipengaruhi oleh variabel yang tidak dimasukkan ke dalam model penelitian.

### 3.7 Implementasi Antarmuka

#### 1) Tampilan Halaman Utama



**Machine Learning Untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linier Berganda**

Tanggal :  
09/07/2020

Komoditas :  
Susu Bubuk Dancow 400kg

Pasar :  
 Jatinegara  
 Kramatjati  
 Minggu

**Prediksi** **Hapus**

**Gambar 17. Tampilan Halaman Utama**

## 2) Tampilan Halaman Hasil Prediksi



Gambar 18. Tampilan Halaman Hasil Prediksi

## IV. PENUTUP

Berdasarkan analisis, perancangan, implementasi dan pengujian dari sistem *machine learning* untuk model prediksi harga sembako dengan metode regresi linear berganda menggunakan *sample data* sembako di DKI Jakarta, maka diperoleh beberapa kesimpulan sebagai berikut:

- 1) Tidak terjadinya pelanggaran asumsi klasik pada himpunan data meliputi linieritas dalam variabel, normalitas residual, homoskedastisitas, non-multikolinieritas, dan non-autokorelasi.
- 2) Hasil uji hipotesis menggunakan uji parsial menghasilkan variabel tanggal ( $X_1$ ), komoditas ( $X_2$ ) dan pasar ( $X_3$ ) secara parsial berpengaruh secara signifikan terhadap variabel harga ( $Y$ ).
- 3) Hasil uji simultan model regresi menerangkan variabel bebas ( $X_1$ ,  $X_2$ , dan  $X_3$ ) yaitu tanggal, komoditas, dan pasar mempengaruhi secara signifikan terhadap variabel terikat ( $Y$ ) yaitu harga.
- 4) Persentase sumbangan pengaruh variabel bebas terhadap variabel terikat sebesar 84,2%, sedangkan sisanya sebesar 15,8% dipengaruhi oleh variabel yang tidak dimasukkan ke dalam penelitian ini.
- 5) Sistem *machine learning* dapat digunakan sebagai alat bantu untuk memprediksi harga sembako harian baik harga dimasa lampau, masa sekarang maupun masa yang akan datang.
- 6) Untuk pembangunan selanjutnya, memperbanyak himpunan data untuk proses pelatihan, sehingga memiliki lebih banyak data untuk dipelajari yang secara langsung berpengaruh besar pada akurasi dari model yang dihasilkan.

- 7) Untuk pembangunan selanjutnya, diharapkan sistem dapat digunakan daerah lain yang membutuhkan *machine learning* untuk memprediksi harga sembako harian dengan menggunakan *sample* data dari daerah setempat.

#### DAFTAR PUSTAKA

- Arthur, L Samuel. (1959). *Some Studies in Machine Learning Using the Game of Checkers*, IBM Journal.
- Brownlee, Jasson. (2015). *Basic Concepts in Machine Learning; Machine Learning Mastery*, <https://machinelearningmastery.com>, diakses tanggal 27 Februari 2020.
- Chollet, Francois. (2018). *Deep Learning with Python*, Manning Publications Co, New York.
- Gujarati, Damodar N. (2003). *Basic Econometrics*, Ed.4, McGraw-Hill Companies Inc, New York.
- Han, Jiawei., Kamber, Micheline., Pei, Jian. (2012). *Data Mining Concepts and Techniques*, Ed.3, Elsevier Inc, USA.
- Harrington, Peter. (2012). *Machine Learning in Action*, Manning Publications Co, New York.
- Kurniawan, D. (2008). *Regresi Linier*, R Development Core Team, Vienna.
- Mitchell, Tom M. (1997). *Machine Learning*, McGraw-Hill Companies Inc, New York.
- Nafi'iyah, Nur. (2015). Seminar Nasional Pengembangan Aktual Teknologi Informasi (SENA BAKTI), *Penerapan Regresi Linear dalam Memprediksi Harga Jual Mobil Bekas*, Lamongan, 2 Desember.
- Peraturan Menteri Perdagangan Republik Indonesia Nomor 07 Tahun 2020 *Tentang Harga Acuan Pembelian di Tingkat Petani dan Harga Acuan Penjualan di Tingkat Konsumen*, Sekretariat Jenderal Kementerian Perdagangan, Jakarta.
- Schmidheiny, Kurt. (2019). *The Multiple Linear Regression Model*. Universitas Basel, Basel.
- Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. Informatika, Bandung.
- Wahyono, Teguh. (2018). *Fundamental Of Python For Machine Learning (Dasar-dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan)*. Gava Media, Yogyakarta.