

Final Project

Anggota kelompok:

No	Nama	NPM	Kontribusi	Tingkat kontribusi
1	Azarine Aisyah	2206051550	Terlibat aktif dalam diskusi dan preprocessing data, menulis laporan.	100%
2	Farah Khairunnisa Prakasa	2206829396	Terlibat aktif dalam diskusi, mencari data, mengerjakan preprocessing data, menulis laporan.	100%
3	Nadia Erlangga	2206051361	Terlibat aktif dalam diskusi, mencari data, mengerjakan coding R dan python, menulis laporan.	100%
4	Nadira Eka Rahmaharva	2206051525	Terlibat aktif dalam diskusi, mencari data, mengerjakan pengolahan data, menulis laporan.	100%
5	Yiesha Reyhani Ghozali	2206828115	Terlibat aktif dalam diskusi, mencari data, mengerjakan preprocessing data, menulis laporan.	100%

Bagian 1. Pendahuluan

Di suatu industri atau bisnis, kepuasan pelanggan seringkali menjadi tujuan utama. Akan tetapi, terdapat sejumlah faktor yang dapat memengaruhi tingkat kepuasan pelanggan tersebut. Ada beberapa faktor yang dapat memengaruhi tingkat kepuasan pelanggan di toko, termasuk Price (harga), Quality (kualitas), Trust (kepercayaan), Customer Satisfaction (kepuasan pelanggan), dan Negative Publicity (publisitas negatif).

Price (harga) seringkali menjadi faktor penentu dalam pengalaman belanja. Harga yang terlalu tinggi dapat membuat pelanggan mencari opsi lain, sementara harga yang terlalu rendah mungkin membuat mereka meragukan kualitas produk atau layanan yang ditawarkan. Quality (kualitas) dari produk atau layanan yang disediakan oleh toko juga memainkan peran penting. Pelanggan mencari produk atau layanan yang memenuhi atau melebihi harapan mereka. Kualitas yang rendah dapat menurunkan kepuasan pelanggan.

Trust (kepercayaan) juga merupakan faktor kunci. Pelanggan harus merasa bahwa toko dapat dipercaya dalam hal produk, layanan, dan kebijakan mereka. Jika ada keraguan terhadap kepercayaan, hal ini dapat merusak hubungan toko dengan pelanggan. Customer Satisfaction (kepuasan pelanggan) adalah hasil dari interaksi antara pelanggan dan semua faktor-faktor di atas. Kepuasan pelanggan mencerminkan sejauh mana toko telah memenuhi atau melebihi harapan pelanggan. Negative Publicity (publisitas negatif) dapat sangat merugikan. Berita buruk atau kontroversi seputar toko dapat merusak citra dan kepercayaan pelanggan terhadap toko tersebut.

Pada kesempatan kali ini, kami akan membahas faktor-faktor yang memengaruhi kepuasan pelanggan. Akan dicari variabel-variabel yang mempengaruhi kepuasan pelanggan dan seberapa signifikan pengaruhnya terhadap tingkat kesetiaan pada toko. Data diambil dari Consumer loyalty In Retail dari data.world.

Dataset terdiri dari 1712 baris dan 6 kolom.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1712 entries, 0 to 1711
Data columns (total 6 columns):
#   Column                      Non-Null Count  Dtype  
---  -
0   Loyalty                     1712 non-null   float64
1   Price                       1712 non-null   float64
2   Quality                     1712 non-null   float64
3   Trust                       1712 non-null   float64
4   Customer satisfaction       1712 non-null   float64
5   Negative publicity          1712 non-null   float64
dtypes: float64(6)
memory usage: 80.4 KB
```

Berikut adalah variabel dan tipe data yang ada, antara lain:

Variabel prediktor:

- Price = kepuasan pelanggan terhadap harga produk, tipe datanya float64
- Quality = kepuasan pelanggan terhadap kualitas produk, tipe datanya float64
- Trust = tingkat kepercayaan pelanggan pada toko, tipe datanya float64
- Customer Satisfaction = kepuasan pelanggan terhadap pelayanan toko, tipe datanya float64
- Negative Publicity = tingkat citra toko di mata publik, tipe datanya float64

Variabel respons/target:

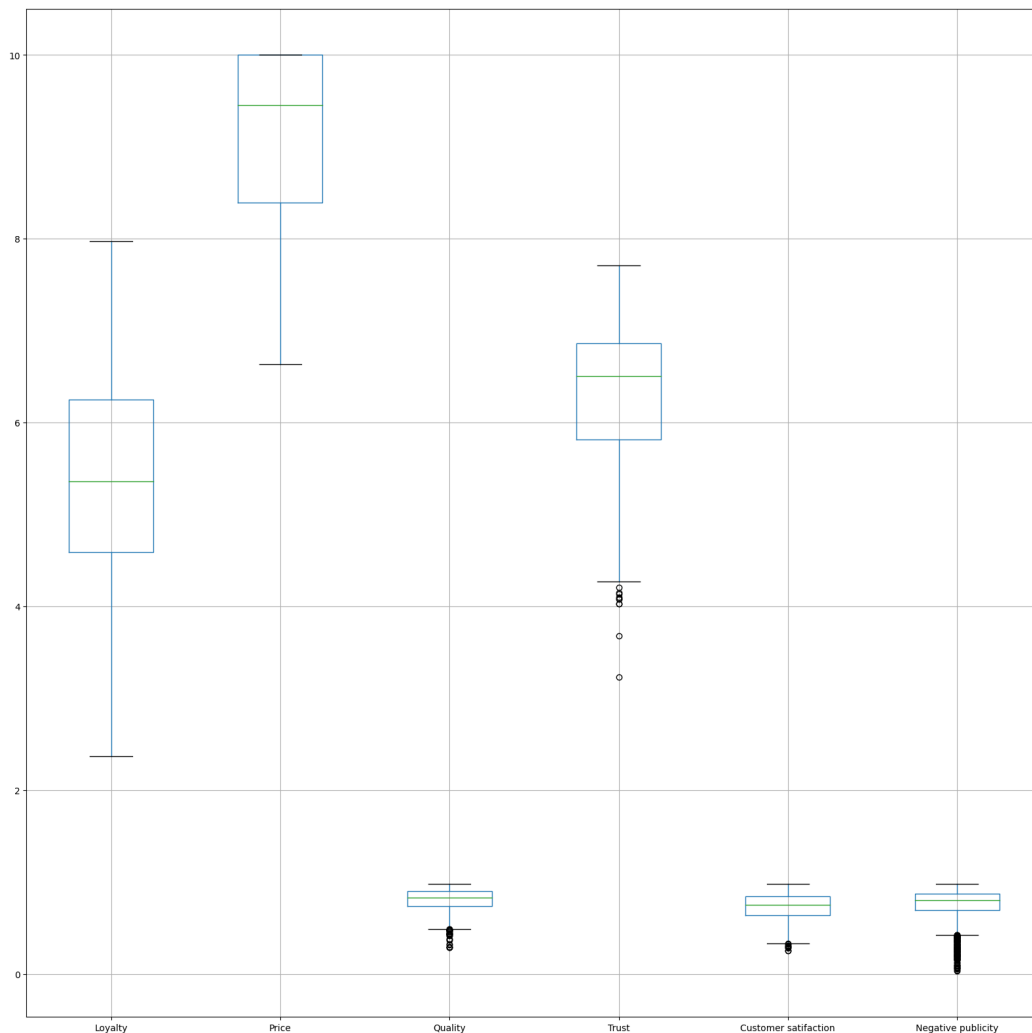
- Loyalty = kesetiaan pelanggan terhadap toko, tipe datanya float64

Seluruh variabel merupakan variabel kualitatif.

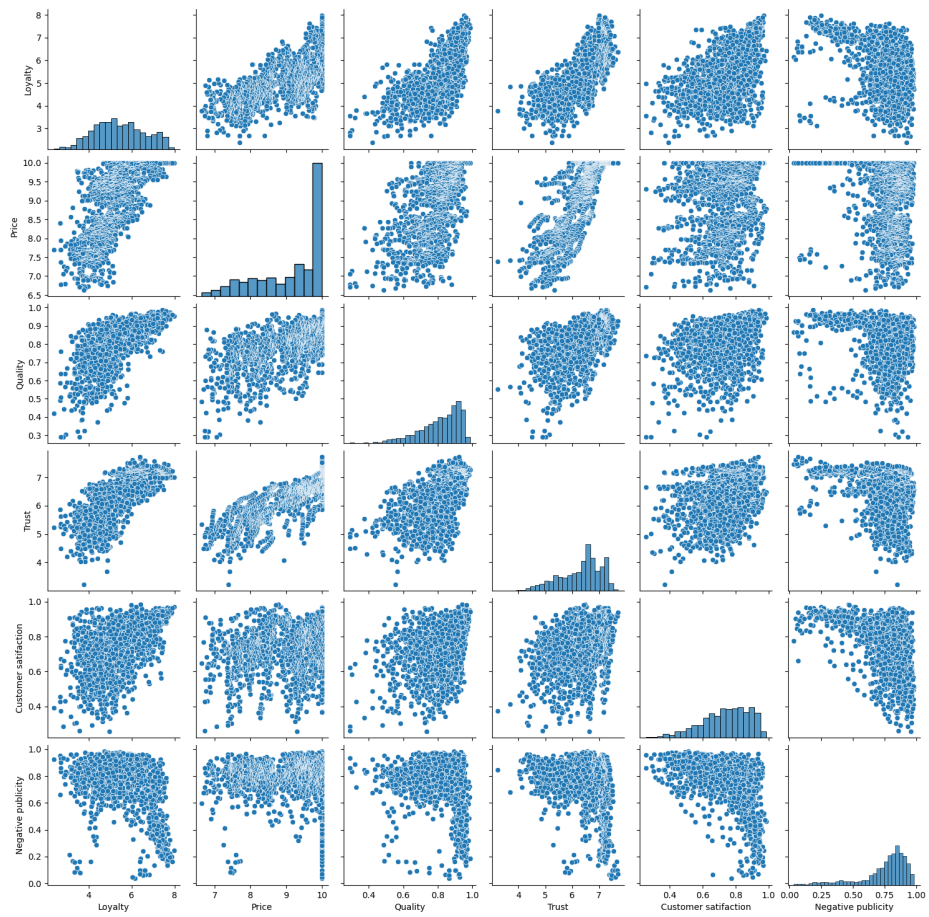
Bagian 2. Pre-processing dan analisis deskriptif

Data source atau sumber data yang digunakan merupakan data sekunder atau data yang telah tersedia. Data yang ada termasuk data observasi karena diperoleh berdasarkan survei. Sebagian besar *pre-processing* dikerjakan menggunakan *python*.

Pertama-tama, kami meng-*import* modul yang dibutuhkan dan juga meng-*import* data. Dataset yang digunakan tidak memiliki *missing value*. Kami membuat box-plot untuk melihat *outliers* pada data. Nilai-nilai yang dianggap *outliers* tidak jauh berbeda dengan nilai-nilai lain pada data, masing-masing sehingga kami beranggapan bahwa nilai tersebut tidak perlu dihilangkan. Kemudian kami mencari plot dari variabel untuk melihat korelasinya menggunakan *pair plot* dan *heatmap*.



[Box Plot]



[Pair Plot]



[Heatmap]

Setelah selesai melakukan *pre-processing*, kami menyimpan hasilnya untuk digunakan pada proses selanjutnya. Sesuai hasil eksplorasi dan visualisasi kami pada tahap *pre-processing*, kami membuat hipotesis yang akan dicek dalam melakukan regresi.

Hipotesis awal yang kami buat adalah tidak terdapat faktor yang mempengaruhi tingkat kepuasan pelanggan terhadap toko, yang artinya parameter bernilai nol. Kemudian, hipotesis alternatifnya adalah terdapat setidaknya satu faktor yang mempengaruhi tingkat kepuasan pelanggan terhadap toko. Hipotesis dapat dinotasikan sebagai berikut.

Hipotesis:

$$H_0 : \beta_1 = \dots = \beta_j = 0, j = 1, 2, \dots, 5$$

$$H_1 : \text{terdapat setidaknya satu } \beta_j \neq 0, j = 1, 2, \dots, 5$$

Untuk melakukan *model validation*, kami membagi data menjadi perbandingan 70:30. 70 digunakan untuk membuat pemodelan dan 30 akan kami gunakan untuk menguji model tersebut.

Bagian 3. Pemodelan

Kami melakukan regresi berganda dengan variabel prediktor adalah semua variabel, yaitu 'Price', 'Quality', 'Trust', 'Customer Satisfaction', serta 'Negative Publicity' dan variabel target adalah 'Loyalty'.

Asumsi-asumsi model:

1. Hubungan antar variabel linier.
2. Error berdistribusi normal.
3. Homoskedastisitas error.
4. Observasi yang independen satu dengan lainnya.

Model regresi:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$$

Keterangan:

y_i = variabel respons Loyalty

β_0 = konstanta / intersep (prediksi nilai saat variabel bebas tidak memengaruhi)

β_1	= parameter Price	x_{i2}	= variabel prediktor Quality
β_2	= parameter Quality	x_{i3}	= variabel prediktor Trust
β_3	= parameter Trust	x_{i4}	= variabel prediktor Customer Satisfaction
β_4	= parameter Customer Satisfaction	x_{i5}	= variabel prediktor Negative Publicity
β_5	= parameter Negative Publicity	Publicity	
x_{i1}	= variabel prediktor Price	ϵ_i	= error

Alasan kami mengajukan model ini adalah karena pada *heatmap* terlihat bahwa semua variabel yaitu Price, Quality, Trust, Customer Satisfaction, dan Negative Publicity memiliki korelasi yang tinggi dengan variabel respon yaitu Loyalty. Selain itu, pada *scatter plot* dalam gambar *pair plot* dapat dilihat bahwa semua variabel memiliki hubungan yang linier dengan Loyalty.

Bagian 4. Pengolahan data dan analisis hasil

Data hasil *pre-processing* menyisakan 1712 unit observasi untuk 6 variabel yang akan digunakan.

Uji F-Global

Akan dianalisis model yang kami ajukan pada bagian sebelumnya dengan melakukan uji F-Global untuk melihat variabel yang memberikan pengaruh signifikan terhadap respon.

- Hipotesis

$$H_0 : \beta_1 = \dots = \beta_j = 0, j = 1, 2, \dots, 5$$

$$H_1 : \text{terdapat setidaknya satu } \beta_j \neq 0, j = 1, 2, \dots, 5$$

- Tingkat Signifikansi

$$\alpha = 0.05$$

- Statistik Uji

```
Call:
lm(formula = Loyalty ~ Price + Quality + Trust + Satisfaction +
    Negative_publicity, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-1.99759 -0.35885  0.01297  0.37683  1.89767

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.0400    0.2215  -9.208  < 2e-16 ***
Price           0.2658    0.0367   7.241 7.98e-13 ***
Quality        2.7890    0.2031  13.736  < 2e-16 ***
Trust          0.4013    0.0435   9.226  < 2e-16 ***
Satisfaction   1.3615    0.1493   9.117  < 2e-16 ***
Negative_publicity -0.9834    0.1092  -9.004  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5873 on 1194 degrees of freedom
Multiple R-squared:  0.7389,    Adjusted R-squared:  0.7378
F-statistic: 675.8 on 5 and 1194 DF,  p-value: < 2.2e-16
```

- Aturan Keputusan dan Kesimpulan

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena $p\text{-value} = 2.2e-16 < \alpha = 0.05$, maka H_0 ditolak. Jadi, terdapat cukup bukti untuk mengatakan bahwa setidaknya terdapat 1 variabel yang signifikan.

Didapat pula nilai $R^2_\alpha = 0.7378$. Nilai R^2_α dikatakan tinggi jika mendekati angka 1 sehingga model ini dianggap sudah cukup baik.

Variable Selection

Variable selection dilakukan dengan menggunakan metode *stepwise* sehingga kami melakukan uji t pada model regresi.

```
Call:
lm(formula = Loyalty ~ Price + Quality + Trust + Satisfaction +
    Negative_publicity, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-1.99759 -0.35885  0.01297  0.37683  1.89767

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.0400     0.2215   -9.208 < 2e-16 ***
Price           0.2658     0.0367    7.241 7.98e-13 ***
Quality        2.7890     0.2031   13.736 < 2e-16 ***
Trust          0.4013     0.0435    9.226 < 2e-16 ***
Satisfaction   1.3615     0.1493    9.117 < 2e-16 ***
Negative_publicity -0.9834    0.1092   -9.004 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5873 on 1194 degrees of freedom
Multiple R-squared:  0.7389,    Adjusted R-squared:  0.7378
F-statistic: 675.8 on 5 and 1194 DF,  p-value: < 2.2e-16
```

- Pertama, akan dimasukkan variabel dengan $|t\text{-value}|$ terbesar di antara variabel prediktor yang lain, yaitu Quality.

Didapatkan $R^2 = 0.5239$ dan $R^2_\alpha = 0.5235$. Lalu, kita akan memeriksa C_p , AIC, SBIC.

Diperoleh hasil sebagai berikut:

```
Call:
lm(formula = Loyalty ~ Quality, data = dftrain)

Residuals:
    Min       1Q   Median       3Q      Max
-2.4085 -0.5424 -0.0096  0.6229  2.4580

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.06616     0.15342   -0.431  0.666
Quality       6.80780     0.18751   36.307 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7917 on 1198 degrees of freedom
Multiple R-squared:  0.5239,    Adjusted R-squared:  0.5235
F-statistic: 1318 on 1 and 1198 DF,  p-value: < 2.2e-16
```

$C_p = 981.2777$
 $AIC = 2849.024$
 $SBIC = -558.6268$

```
> ols_mallows_cp(model1, model)
[1] 981.2777
> AIC(model1)
[1] 2849.024
> ols_sbic(model1, model)
[1] -558.6268
```

Model dikatakan baik jika nilai R^2_α tinggi mendekati angka 1. Didapat nilai R^2 dan R^2_α masih sangat kecil serta C_p , AIC, dan SBIC masih sangat besar (semakin kecil nilainya maka semakin baik), maka akan dilanjutkan metodenya.

- Selanjutnya, akan dimasukkan variabel dengan $|t\text{-value}|$ terbesar selanjutnya, yaitu Trust.

Didapatkan $R^2 = 0.6751$ dan $R^2_\alpha = 0.6745$. Lalu, kita akan memeriksa C_p , AIC, SBIC.

Diperoleh hasil sebagai berikut:

- $C_p = 291.8724$

- $AIC = 2392.536$

- $SBIC = -1014.145$

```
lm(formula = Loyalty ~ Quality + Trust, data = dftrain)

Residuals:
    Min       1Q   Median       3Q      Max
-2.03967 -0.43538 -0.01644  0.46377  2.21931

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.38241    0.16034   -14.86  <2e-16 ***
Quality      3.87957    0.19852   19.54  <2e-16 ***
Trust        0.74206    0.03144    23.60  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6543 on 1197 degrees of freedom
Multiple R-squared:  0.6751,    Adjusted R-squared:  0.6745
F-statistic: 1243 on 2 and 1197 DF,  p-value: < 2.2e-16

> ols_mallows_cp(model2, model)
[1] 291.8724
> AIC(model2)
[1] 2392.536
> ols_sbic(model2, model)
[1] -1014.145
```

Model dikatakan baik jika nilai R^2_α tinggi mendekati angka 1. Didapat nilai R^2 dan R^2_α masih kecil serta C_p , AIC, dan SBIC masih sangat besar (semakin kecil nilainya maka semakin baik), maka akan dilanjutkan metodenya.

- Akan dimasukkan variabel selanjutnya, yaitu Negative_publicity.

Didapatkan $R^2 = 0.7117$ dan $R^2_\alpha = 0.7109$. Lalu, kita akan memeriksa C_p , AIC, SBIC.

Diperoleh hasil sebagai berikut:

- $C_p = 126.5259$

- $AIC = 2251.151$

- $SBIC = -1155.037$

```
lm(formula = Loyalty ~ Quality + Trust + Negative_publicity,
    data = dftrain)

Residuals:
    Min       1Q   Median       3Q      Max
-2.09784 -0.38602  0.01527  0.41993  2.10920

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.77852    0.19945   -3.903  1e-04 ***
Quality      3.82221    0.18714   20.424  <2e-16 ***
Trust        0.64558    0.03065    21.064  <2e-16 ***
Negative_publicity -1.26050    0.10231  -12.321  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6166 on 1196 degrees of freedom
Multiple R-squared:  0.7117,    Adjusted R-squared:  0.7109
F-statistic: 984 on 3 and 1196 DF,  p-value: < 2.2e-16

> ols_mallows_cp(model3, model)
[1] 126.5259
> AIC(model3)
[1] 2251.151
> ols_sbic(model3, model)
[1] -1155.037
```

Model dikatakan baik jika nilai R^2_α tinggi mendekati angka 1. Didapat nilai R^2 dan R^2_α masih kecil serta C_p , AIC, dan SBIC masih cukup besar (semakin kecil nilainya maka semakin baik), maka akan dilanjutkan metodenya.

- Lalu dimasukkan variabel selanjutnya, yaitu Satisfaction.

Didapatkan $R^2 = 0.7274$ dan $R^2_\alpha = 0.7265$. Lalu, kita akan memeriksa C_p , AIC, SBIC.

Diperoleh hasil sebagai berikut:

- $C_p = 56.42594$

- $AIC = 2185.67$

- $SBIC = -1220.158$

```
lm(formula = Loyalty ~ Quality + Trust + Negative_publicity +
    Satisfaction, data = dftrain)

Residuals:
    Min       1Q   Median       3Q      Max
-2.1446 -0.3503  0.0111  0.3786  2.0032

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.59068    0.21721  -7.323 4.43e-13 ***
Quality      3.38845    0.18935   17.895  < 2e-16 ***
Trust        0.63469    0.02984    21.269  < 2e-16 ***
Negative_publicity -0.86523    0.11029  -7.845 9.53e-15 ***
Satisfaction  1.26264    0.15187    8.314 2.49e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5998 on 1195 degrees of freedom
Multiple R-squared:  0.7274,    Adjusted R-squared:  0.7265
F-statistic: 797.3 on 4 and 1195 DF,  p-value: < 2.2e-16

> ols_mallows_cp(model4, model)
[1] 56.42594
> AIC(model4)
[1] 2185.67
> ols_sbic(model4, model)
[1] -1220.158
```


Model dikatakan baik jika nilai R^2_α tinggi mendekati angka 1. Didapat nilai R^2 dan R^2_α sudah cukup besar, tetapi C_p , AIC, dan SBIC masih cukup besar (semakin kecil nilainya maka semakin baik), maka akan dilanjutkan metodenya.

- Terakhir, akan dimasukkan variabel Price.

Didapatkan $R^2 = 0.7389$ dan $R^2_\alpha = 0.7378$. Lalu, kita akan memeriksa C_p , AIC, SBIC.

Diperoleh hasil sebagai berikut:

- $C_p = 6$
- AIC = 2136.105
- SBIC = -1269.288

```
> ols_mallows_cp(model5, model)
[1] 6
> AIC(model5)
[1] 2136.105
> ols_sbic(model5, model)
[1] -1269.288
```

```
lm(formula = Loyalty ~ Quality + Trust + Satisfaction + Negative_publicity +
    Price, data = dftrain)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.99759 -0.35885  0.01297  0.37683  1.89767
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.0400     0.2215   -9.208 < 2e-16 ***
Quality        2.7890     0.2031   13.736 < 2e-16 ***
Trust          0.4013     0.0435    9.226 < 2e-16 ***
Satisfaction   1.3615     0.1493    9.117 < 2e-16 ***
Negative_publicity -0.9834    0.1092   -9.004 < 2e-16 ***
Price          0.2658     0.0367    7.241 7.98e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.5873 on 1194 degrees of freedom
Multiple R-squared:  0.7389,    Adjusted R-squared:  0.7378
F-statistic: 675.8 on 5 and 1194 DF,  p-value: < 2.2e-16
```

Model dikatakan baik jika nilai R^2_α tinggi mendekati angka 1. Didapat nilai R^2 dan R^2_α sudah cukup besar serta C_p , AIC, dan SBIC juga sudah cukup kecil dibandingkan sebelumnya. Sehingga, dapat dikatakan bahwa model sudah cukup baik.

Dari metode *stepwise* didapatkan **pemodelan terbaik** adalah

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$$

Keterangan:

y_i = variabel respons Loyalty

β_0 = konstanta / intersep (prediksi nilai saat variabel bebas tidak memengaruhi)

β_1 = parameter Price

x_{i2} = variabel prediktor Quality

β_2 = parameter Quality

x_{i3} = variabel prediktor Trust

β_3 = parameter Trust

x_{i4} = variabel prediktor Customer

β_4 = parameter Customer Satisfaction

Satisfaction

β_5 = parameter Negative Publicity

x_{i5} = variabel prediktor Negative

Publicity

x_{i1} = variabel prediktor Price

ε_i = error

Variance Inflation Factor (VIF)

Price	Quality	Trust	Customer.satisfaction	Negative.publicity
4.243221	2.131259	3.899322	1.540964	1.418525

Karena semua VIF < 10 maka dapat kita simpulkan bahwa tidak terdapat multikolinearitas.

Uji Signifikansi Interaksi

Berdasarkan *heatmap* pada Bagian 2, kami mendapati bahwa ada korelasi linier yang cukup tinggi antara variabel Price dan Trust dengan korelasi 0.62. Kemudian, kami mencoba untuk membuat model yang ditambahkan interaksi antara Price dan Trust dari pemodelan terbaik yang diperoleh dengan metode *stepwise*.

- Full Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i1} x_{i3} + \varepsilon_i$$

- Reduced Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$$

- Hipotesis

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0$$

- Tingkat Signifikansi

$$\alpha = 0.05$$

- Statistik Uji

Analysis of Variance Table

```

Model 1: Loyalty ~ Price + Quality + Trust + Satisfaction + Negative_publicity
Model 2: Loyalty ~ Price + Quality + Trust + Satisfaction + Negative_publicity +
  Price * Trust
  Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1    1194 411.83
2    1193 378.19  1    33.644 106.13 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

- Aturan Keputusan dan Kesimpulan

Didapatkan p-value = $2.2e-16 < \alpha = 0.05$ sehingga H_0 ditolak yang artinya interaksi memiliki pengaruh yang signifikan pada respons.

Model Validation

Pada langkah ini kita akan mengecek dua jenis model, yaitu model dengan interaksi dan model yang tidak ada interaksi pada data baru yang telah di-*split* sebelumnya

Model tanpa interaksi

R2	RMSE	MAE
0.7162578	0.5925223	0.4547313

Model dengan interaksi

R2	RMSE	MAE
0.7410832	0.565884	0.4393578

Dari hasil validasi tersebut, dapat dilihat bahwa model dengan interaksi memiliki RMSE dan MAE lebih kecil dibandingkan dengan model yang tidak memiliki interaksi. Semakin kecil nilai RMSE dan MAE, maka semakin baik model tersebut.

Hal ini juga dapat dilihat dari AIC model dengan interaksi = 2074.248 < AIC model tanpa interaksi = 2172.0984, begitu pula SBIC model dengan interaksi = -1330.373 < SBIC model tanpa interaksi = -1233.2938.

Maka, dapat disimpulkan bahwa model terbaiknya adalah model dengan interaksi.

Residual Analysis

Kami melakukan *residual analysis* untuk memastikan apakah model yang dipilih sudah tepat atau belum. Dengan melihat persebaran residual variabel respons, akan diperiksa apakah model memenuhi asumsi independensi, linearitas, normalitas, dan variansi konstan (homoskedastisitas).

1. Uji Asumsi Independensi Residual

Kami menggunakan uji Durbin Watson untuk memeriksa apakah residualnya independen.

Hipotesis:

H_0 : Tidak ada autokorelasi

H_1 : Ada autokorelasi

Tingkat Signifikansi:

$$\alpha = 0.05$$

Statistik Uji:

Durbin-Watson test

```
data: model_int
DW = 1.8224, p-value = 0.0008239
alternative hypothesis: true autocorrelation is greater than 0
```

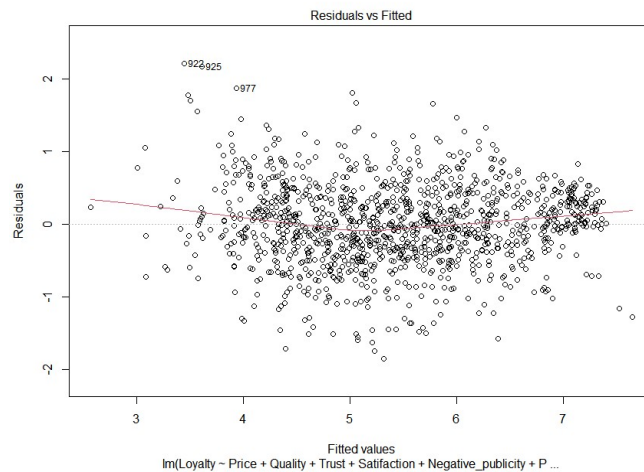
Aturan keputusan:

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena $p\text{-value} = 0.0008239 < 0.05 = \alpha$, maka H_0 ditolak. Sehingga dapat disimpulkan bahwa residual tidak memenuhi asumsi independensi.

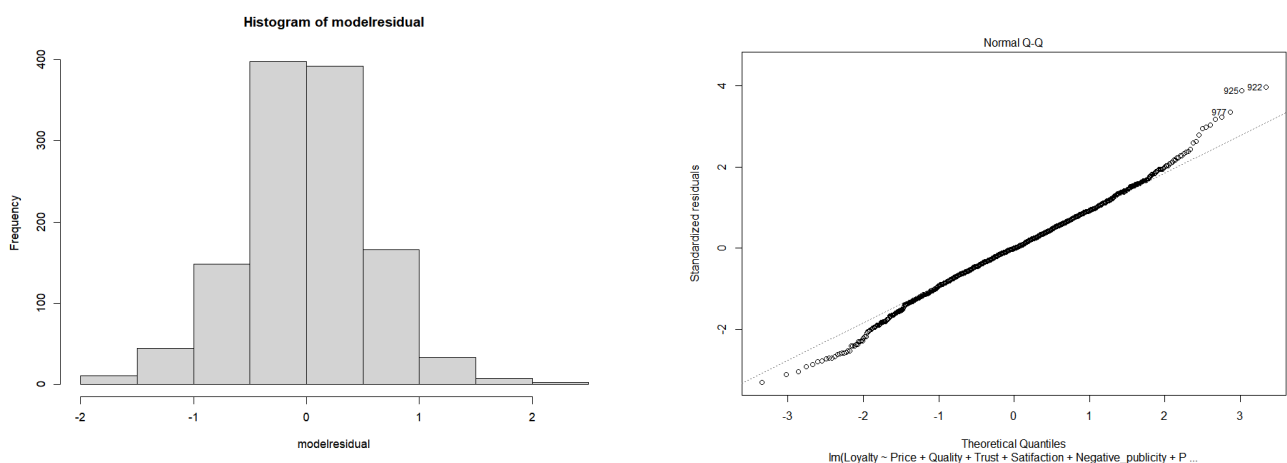
2. Uji Asumsi Linearitas Residual

Plot ini menunjukkan sebaran residu terhadap nilai-nilai yang diprediksi oleh model.



3. Uji Asumsi Normalitas Residual

Normal Q-Q (quantile-quantile) Plot dan Histogram



Histogram dan Quantile-Quantile (Q-Q) plot dari residu digunakan untuk mengevaluasi distribusi residu. Pada plot histogram terlihat menyerupai lonceng dan Q-Q plot cukup mengikuti garis diagonal. Oleh karena itu, dapat dikatakan bahwa residu berdistribusi normal.

4. Uji Asumsi Variansi Konstan (Homoskedastisitas) Residual

Kami menggunakan uji *studentized* Breusch-Pagan untuk memeriksa apakah variabilitas residual konstan.

Hipotesis:

H_0 : Varians residual bersifat homoskedastisitas atau $\text{var}(\varepsilon_i) = \sigma^2$

H_1 : Varians residual bersifat heteroskedastisitas atau $\text{var}(\varepsilon_i) \neq \sigma^2$

Tingkat Signifikansi:

$$\alpha = 0.05$$

Statistik Uji:

studentized Breusch-Pagan test

```
data: model_int  
BP = 61.002, df = 6, p-value = 2.816e-11
```

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena $p\text{-value} = 2.816e-11 < 0.05 = \alpha$, maka H_0 ditolak. Sehingga, dapat disimpulkan bahwa residual tidak memenuhi asumsi homoskedastisitas.

Dari uji asumsi residual yang telah dilakukan, asumsi independen dan asumsi homoskedastisitas tidak terpenuhi. Ada kemungkinan bahwa hasil yang diperoleh bias sehingga kesimpulan mungkin tidak *reliable*. Hal yang dapat dilakukan untuk mengatasi hal tersebut adalah menambahkan sampel agar variansi lebih stabil atau regresi dapat dilakukan dengan metode non-parametrik.

Bagian 5. Penutup

Berdasarkan model yang kami ajukan dan setelah dilakukan pengolahan data, dapat kami ambil kesimpulan bahwa **model terbaik** adalah:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i1} x_{i3} + \varepsilon_i$$

Model tersebut adalah model terbaik karena semua variabel yang dicantumkan signifikan pengaruhnya terhadap variabel respons, dengan nilai R^2_α yang cukup tinggi yakni $R^2_\alpha = 0.7492$.

Artinya, kepuasan harga yang ditawarkan, kualitas produk, kepercayaan pelanggan, pelayanan yang diberikan, dan citra toko, serta interaksi antara harga yang ditawarkan dengan kepercayaan pelanggan mempengaruhi loyalitas pelanggan terhadap toko secara signifikan. Model ini mencantumkan interaksi antara harga yang ditawarkan dengan kepercayaan pelanggan karena tingkat korelasi linier yang cukup tinggi, yakni 0.62. Berdasarkan *heatmap* pada bagian 2, faktor yang memiliki korelasi paling tinggi dengan respons adalah kepercayaan pelanggan dan kualitas produk, yakni 0.66.

Bagian 6. Lampiran

Link GDrive: [UAS_MODEL LINIER](#)