



Minggu - 6

Fungsi Minor 3: Prediksi

Bayu Widodo

Teknologi Rekayasa Perangkat Lunak, SV-IPB

18/08/2025 (updated: 24 September 2025)



| No | Jarak | Waktu |
|----|-------|-------|
| 1 | 0.50 | 9.95 |
| 2 | 1.10 | 24.45 |
| 3 | 1.20 | 31.75 |
| 4 | 5.50 | 35.00 |
| 5 | 2.95 | 25.02 |
| 6 | 2.00 | 16.86 |
| 7 | 3.75 | 14.38 |
| 8 | 0.52 | 9.60 |
| 9 | 1.00 | 24.35 |
| 10 | 3.00 | 27.50 |
| 11 | 4.12 | 17.08 |
| 12 | 4.00 | 4.00 |
| 13 | 5.00 | 5.00 |
| 14 | 3.60 | 11.66 |

| No | Jarak | Waktu |
|----|-------|-------|
| 15 | 2.05 | 21.65 |
| 16 | 4.00 | 17.89 |
| 17 | 6.00 | 69.00 |
| 18 | 5.85 | 10.30 |
| 19 | 5.40 | 34.93 |
| 20 | 2.50 | 46.59 |
| 21 | 2.90 | 44.88 |
| 22 | 5.10 | 54.12 |
| 23 | 5.90 | 56.23 |
| 24 | 1.00 | 22.13 |
| 25 | 4.00 | 4.00 |

Tabel di atas menunjukkan hubungan antara jarak tempuh dan waktu pengantaran ojek online. Jika pesanan ke-26 memiliki jarak 1,5 km, berapa lama waktu pengantaran yang diperkirakan?



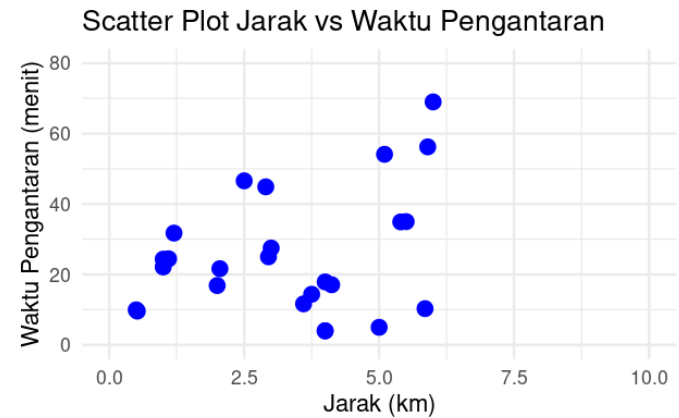
```
if(!require(dplyr)){install.packages("dplyr")}  
library(dplyr)  
  
jarak<- c(0.50,1.10,1.20,5.50,2.95,2.00,  
          3.75,0.52,1.00,3.00,  
          4.12,4.00,5.00,3.60,2.05,4.00,6.00,  
          5.85,5.40,2.50,2.90,5.10,5.90,  
          1.00,4.00)  
waktu <- c(9.95,24.45,31.75,35.00,25.02,16.86,  
           14.38,9.60,24.35,27.50,17.08,  
           4.00,5.00,11.66,21.65,17.89,69.00,  
           10.30,34.93,46.59,44.88,54.12,  
           56.23,22.13,4.00)  
antar_pesan <- data.frame(jarak,waktu)
```

```
ringkasan <- antar_pesan %>%  
  summarise(  
    n = n(), # jumlah observasi  
    mean_jarak = mean(jarak),  
    median_jarak = median(jarak),  
    sd_jarak = sd(jarak),  
    min_jarak = min(jarak),  
    max_jarak = max(jarak),  
    mean_waktu = mean(waktu),  
    median_waktu = median(waktu),  
    sd_waktu = sd(waktu),  
    min_waktu = min(waktu),  
    max_waktu = max(waktu),  
    korelasi = cor(antar_pesan$waktu,  
                   antar_pesan$jarak)  
  )  
str(ringkasan)
```



```
library(ggplot2)

ggplot(antar_pesan, aes(x = jarak, y = waktu)) +
  geom_point(color = "blue", size = 3) +
  labs(
    title = "Scatter Plot Jarak  
vs Waktu Pengantaran",
    x = "Jarak (km)",
    y = "Waktu Pengantaran (menit)"
  ) +
  xlim(0, 10) +
  ylim(0, 80) +
  theme_minimal()
```





```
# Model regresi
model <- lm(waktu ~ jarak, data = antar_pesan)
coef_model <- coef(model)
summary_model <- summary(model)
summary_model
```

1. Persamaan regresi: $\text{waktu} = 15.21 + (3.11 \times \text{jarak})$. Artinya, jika jarak bertambah 1 satuan, waktu cenderung bertambah sekitar 3.11 satuan.
2. Goodness of fit:
 - $R^2 = 0.1032$, hanya sekitar 10.3% variasi waktu yang dapat dijelaskan oleh jarak.
 - Adjusted $R^2 = 0.064$, model kurang baik dalam menjelaskan data. Hanya 6.4% variasi waktu yang benar-benar dijelaskan oleh jarak.
3. Uji F dalam regresi digunakan untuk melihat apakah model secara keseluruhan bermanfaat. Artinya: apakah variabel independen (di sini: jarak) benar-benar bisa menjelaskan variasi variabel dependen (waktu) dibandingkan dengan model tanpa variabel (hanya intercept). Hasil uji:
 - F-statistic = 2.648, artinya model regresi dengan variabel jarak hanya sedikit lebih baik dibandingkan model tanpa variabel.
 - Karena $p > 0.05$, maka model tidak signifikan secara keseluruhan.
 - Artinya jarak tidak terbukti menjelaskan variasi waktu secara nyata.
4. Artinya, masih ada faktor lain selain jarak yang lebih berpengaruh terhadap waktu (misalnya kondisi jalan, kecepatan, atau hambatan lain). Jadi, regresi ini tidak bisa digunakan untuk prediksi yang akurat.



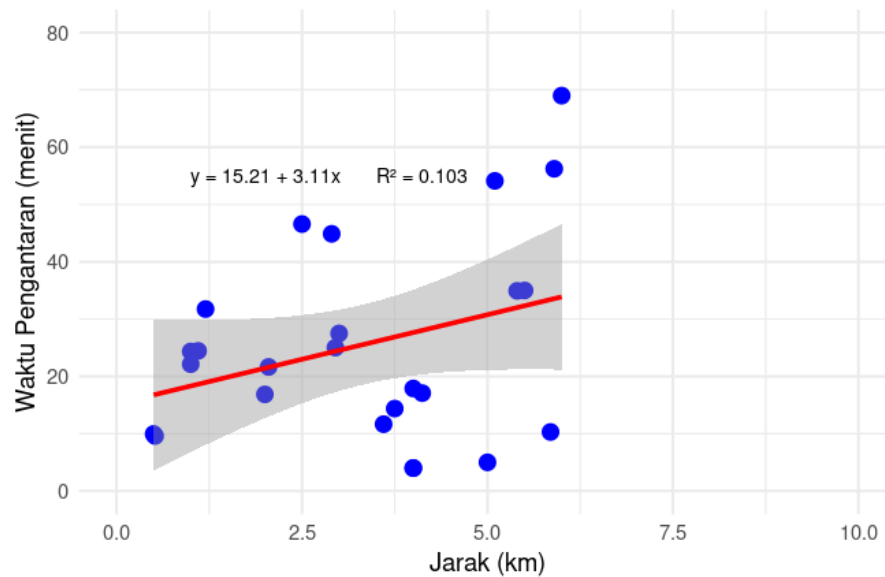
```
# Buat persamaan regresi dalam bentuk teks
persamaan <- paste0(
  "y = ", round(coef_model[1], 2),
  " + ", round(coef_model[2], 2), "x"
)
# Ambil nilai R-squared
r2 <- round(summary_model$r.squared, 3)
teks_r2 <- paste0("R2 = ", r2)

# Scatter + garis regresi + persamaan
ggplot(antar_pesan, aes(x = jarak, y = waktu)) +
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "lm", se = TRUE, color = "red") +
  annotate("text", x = 1, y = 55, label = persamaan,
    color = "black", size = 3, hjust = 0) +
  annotate("text", x = 3.5, y = 55, label = teks_r2,
    color = "black", size = 3, hjust = 0) +

  labs(
    title = "Regresi Linier: Jarak vs Waktu Pengantaran",
    x = "Jarak (km)",
    y = "Waktu Pengantaran (menit)"
  ) +
  xlim(0, 10) +
  ylim(0, 80) +
  theme_minimal()
```



Regresi Linier: Jarak vs Waktu Pengantaran



```
# model regresi
model <- lm(waktu ~ jarak,
            data = antar_pesan
            )
# memprediksi untuk data baru
data_baru <- data.frame(
  jarak = c(5, 10, 15)
)
predict(
  model, newdata = data_baru
)
```



Tugas Praktikum Data Mining Prediction dengan Linear Regression (Dataset: longley)

- Mahasiswa memahami konsep regresi linear dalam prediksi data ekonomi.
- Mahasiswa mampu menggunakan R (base dan atau tidyverse) untuk membangun model prediksi.
- Mahasiswa dapat melakukan interpretasi hasil model serta membuat prediksi untuk data baru.



Petunjuk:

- Gunakan dataset bawaan R: longley
- Berisi data ekonomi Amerika Serikat tahun 1947–1962.
- Variabel utama:
 - GNP : Produk Nasional Bruto (miliar USD)
 - Unemployed : Jumlah pengangguran (ribu orang)
 - Population : Jumlah populasi (juta orang)
 - Employed : Jumlah penduduk bekerja (juta orang) → Target Y

Persiapan Dataset

```
# Load dataset
data(longley)
# Lihat struktur data
str(longley)
# Ringkasan statistik
summary(longley)
```



Tugas Praktikum Analisis Regresi

1. Visualisasi Hubungan Variabel

- Buat visualisasi untuk melihat hubungan antara variabel GNP dengan Employed menggunakan grafik scatter plot.

2. Model Regresi Linear Sederhana

- Bangun model regresi linear sederhana dengan GNP sebagai variabel prediktor dan Employed sebagai variabel respon.

3. Prediksi dengan Regresi Linear Sederhana

- Lakukan prediksi nilai Employed berdasarkan nilai GNP baru: c(400, 500, 600).

4. Model Regresi Linear Berganda

- Bangun model regresi linear berganda dengan Employed sebagai variabel respon dan GNP, Unemployed, serta Population sebagai variabel prediktor.

5. Prediksi dengan Regresi Linear Berganda

- Lakukan prediksi nilai Employed berdasarkan data baru berikut:

```
new_data2 <- data.frame(  
  GNP = c(550, 600),  
  Unemployed = c(250, 200),  
  Population = c(115, 120)  
)
```