Matematika 5A - DS1

ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEPUASAN PENUMPANG PESAWAT

Kelompok 6

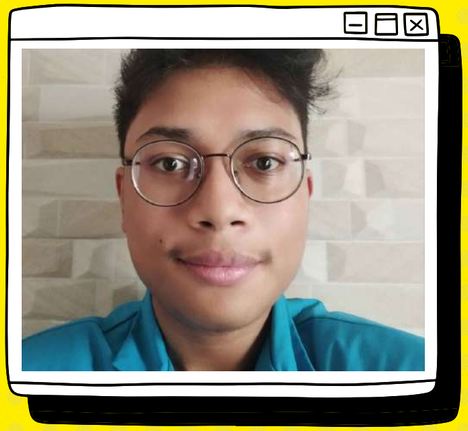
THE TEAM

Dani Hidayat
(11220940000014)

M. Fahmi Islam Fadlillah
(11220940000016)

M. Abdul Ghofur
(11220940000020)

M. Anthar Jailani Putra
(11220940000029)



Fadhilah Awal Rizqiansyah
(11220940000064)

LANGKAH-LANGKAH

EKSPLORASI DATA

MODEL BUILDING

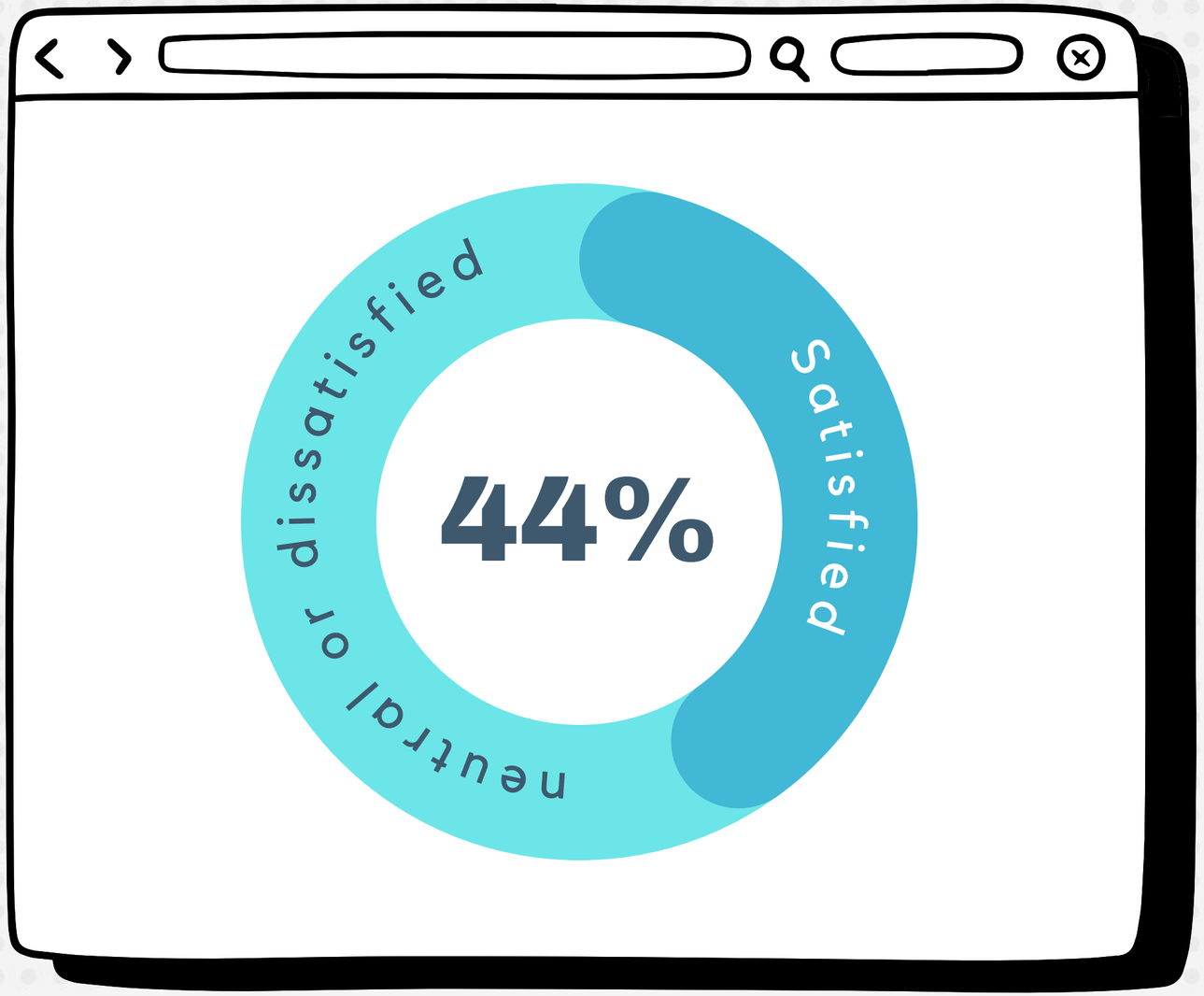
HASIL DAN PEMBAHASAN

RINGKASAN DATA

Dataset ini berisi survei kepuasan penumpang maskapai penerbangan.

- Gender : Jenis kelamin penumpang (Female: Perempuan, Male: Laki-laki).
- Age : Usia aktual dari penumpang.
- Type of Travel : Tujuan perjalanan penumpang (Personal Travel: Perjalanan Pribadi, Business Travel: Perjalanan Bisnis).
- Class : Kelas perjalanan di pesawat (Business: Bisnis, Eco: Ekonomi, Eco Plus: Ekonomi Plus).
- Flight Distance : Jarak penerbangan dari perjalanan ini (dalam satuan mil atau km).
- Inflight Wifi Service : Tingkat kepuasan terhadap layanan WiFi di dalam penerbangan (0: Tidak Berlaku; 1-5: Skala Kepuasan).
- Inflight Entertainment : Tingkat kepuasan terhadap hiburan di dalam penerbangan (1-5: Skala Kepuasan).
- Departure Delay in Minutes : Lama keterlambatan keberangkatan dalam menit.
- Arrival Delay in Minutes : Lama keterlambatan kedatangan dalam menit.
- Satisfaction : Tingkat kepuasan terhadap maskapai penerbangan (Satisfaction: Puas, Neutral or Dissatisfaction: Netral atau Tidak Puas).

VISUALISASI VARIABEL RESPON (SATISFACTION)



```
> str(data)
'data.frame':   25976 obs. of  10 variables:
 $ satisfaction               : chr  "satisfied" "satisfied" "neutral or dissatisfied" "satisfied" ...
 $ Age                        : int  52 36 20 44 49 16 77 43 47 46 ...
 $ Flight_Distance            : int  160 2863 192 3377 1182 311 3987 2556 556 1744 ...
 $ Departure_Delay_in_Minutes: int  50 0 0 0 0 0 0 77 1 28 ...
 $ Arrival_Delay_in_Minutes   : num  44 0 0 6 20 0 0 65 0 14 ...
 $ Gender                     : chr  "Female" "Female" "Male" "Male" ...
 $ Type_of_Travel             : chr  "Business travel" "Business travel" "Business travel" "Business travel" ...
 $ Class                      : chr  "Eco" "Business" "Eco" "Business" ...
 $ Inflight_wifi_service      : int  5 1 2 0 2 3 5 2 5 2 ...
 $ Inflight_entertainment     : int  5 4 2 1 2 5 5 4 5 4 ...
```

RINGKASAN STRUKTUR DATASET

Tipe data masing-masing kolom, banyak data, dan beberapa isi kolom. Dengan masing-masing tipe data dari 10 variabel di atas seperti berikut:

- Data Numerik: Age, Flight Distance, Departure Delay in Minutes, Arrival Delay in Minutes.
- Data Kategorik: Satisfaction(Target), Gender, Type of Travel, Class, Inflight wifi service, Inflight entertainment.

DISTRIBUSI NILAI DARI VARIABEL KATEGORIK

```
> for (kolom in kategorik) {
+   print(table(data[[kolom]]))
+   print(paste("Distribusi nilai pada kolom:", kolom))
+ }

Female   Male
 13172  12804
[1] "Distribusi nilai pada kolom: Gender"

Business travel Personal Travel
          18038             7938
[1] "Distribusi nilai pada kolom: Type_of_Travel"

Business      Eco Eco Plus
   12495    11564     1917
[1] "Distribusi nilai pada kolom: Class"

   0    1    2    3    4    5
 813 4488 6490 6317 4981 2887
[1] "Distribusi nilai pada kolom: Inflight_wifi_service"

   0    1    2    3    4    5
   4 3197 4331 4745 7368 6331
[1] "Distribusi nilai pada kolom: Inflight_entertainment"

neutral or dissatisfied               satisfied
                  14573                   11403
[1] "Distribusi nilai pada kolom: satisfaction"
```

STATISTIKA DESKRIPTIF DARI VARIABEL NUMERIK

```
> lapply(data[numerik], summary)
$Age
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   7.00   27.00   40.00   39.62   51.00   85.00

$Flight_Distance
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     31     414     849    1194    1744    4983

$Departure_Delay_in_Minutes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    0.00   14.31   12.00 1128.00

$Arrival_Delay_in_Minutes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    0.00   14.69   13.00 1115.00
```

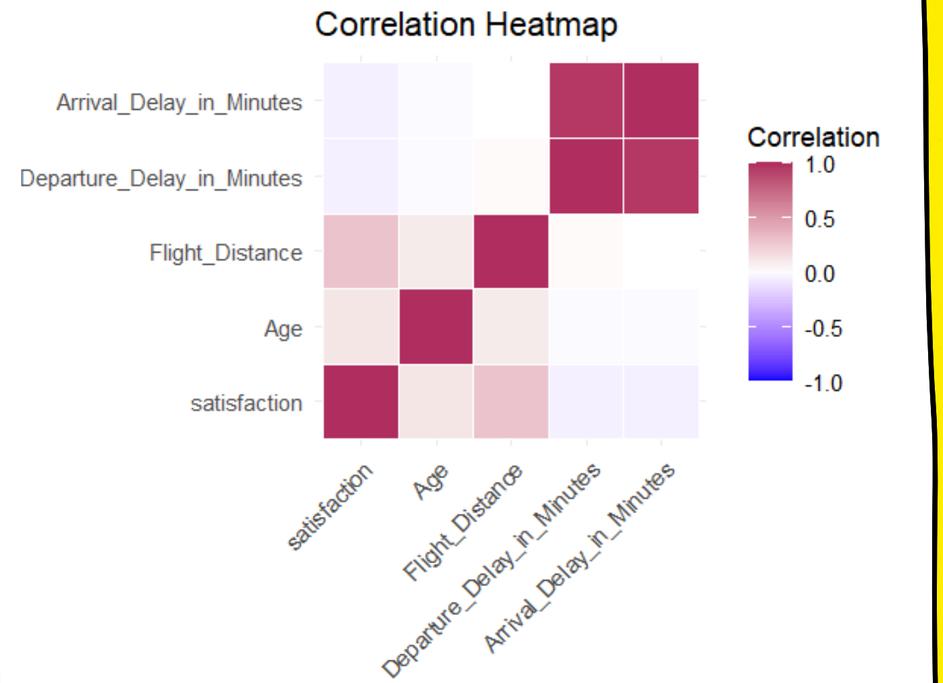
```
> data$satisfaction <- ifelse(data
$satisfaction == "satisfied", 1, 0)
> print(table(data$satisfaction))

    0     1
14573 11403
> data$Inflight_entertainment[data$Inflight_entertainment %in% c(0,1)] <- "Low"
> data$Inflight_entertainment[data$Inflight_entertainment %in% c(2,3)] <- "Medium"
> data$Inflight_entertainment[data$Inflight_entertainment %in% c(4,5)] <- "High"
> data$Inflight_wifi_service[data$Inflight_wifi_service %in% c(0, 1)] <- "Low"
> data$Inflight_wifi_service[data$Inflight_wifi_service %in% c(2,3)] <- "Medium"
> data$Inflight_wifi_service[data$Inflight_wifi_service %in% c(4,5)] <- "High"
> data$Class[data$Class %in% c("Eco Plus")] <- "Eco"

Business      Eco
   12495    13481
[1] "Distribusi nilai pada kolom: Class"

  High    Low Medium
  7868   5301  12807
[1] "Distribusi nilai pada kolom: Inflight_wifi_service"

  High    Low Medium
 13699   3201   9076
[1] "Distribusi nilai pada kolom: Inflight_entertainment"
```

DISTRIBUSI NILAI DARI VARIABEL KATEGORIK YANG DISEDERHANAKAN

VIF MODEL DASAR

```
> model_dasar <- glm(satisfaction ~
+                      Gender +
+                      Age +
+                      Flight_Distance +
+                      Departure_Delay_in_Minutes +
+                      Arrival_Delay_in_Minutes +
+                      Type_of_Travel +
+                      Class +
+                      Inflight_wifi_service +
+                      Inflight_entertainment,
+                    data = data, family = binomial)
> car::vif(model_dasar)
                              GVIF Df GVIF^(1/(2*Df))
Gender                    1.002171  1        1.001085
Age                       1.032955  1        1.016344
Flight_Distance           1.211418  1        1.100644
Departure_Delay_in_Minutes 10.193238  1        3.192685
Arrival_Delay_in_Minutes   10.198556  1        3.193518
Type_of_Travel            1.273969  1        1.128702
Class                     1.617391  1        1.271767
Inflight_wifi_service     1.315644  2        1.070988
Inflight_entertainment    1.092943  2        1.022467
```

KORELASI



MODEL DASAR 2

Dari antar variabel diatas terlihat variabel departure_delay_in_minutes dan arrival_delay_in_minutes mendekati 1.
Variabel arrival_delay_in_minutes kami buang dalam model Dasar, sehingga didapat Model Dasar 2.

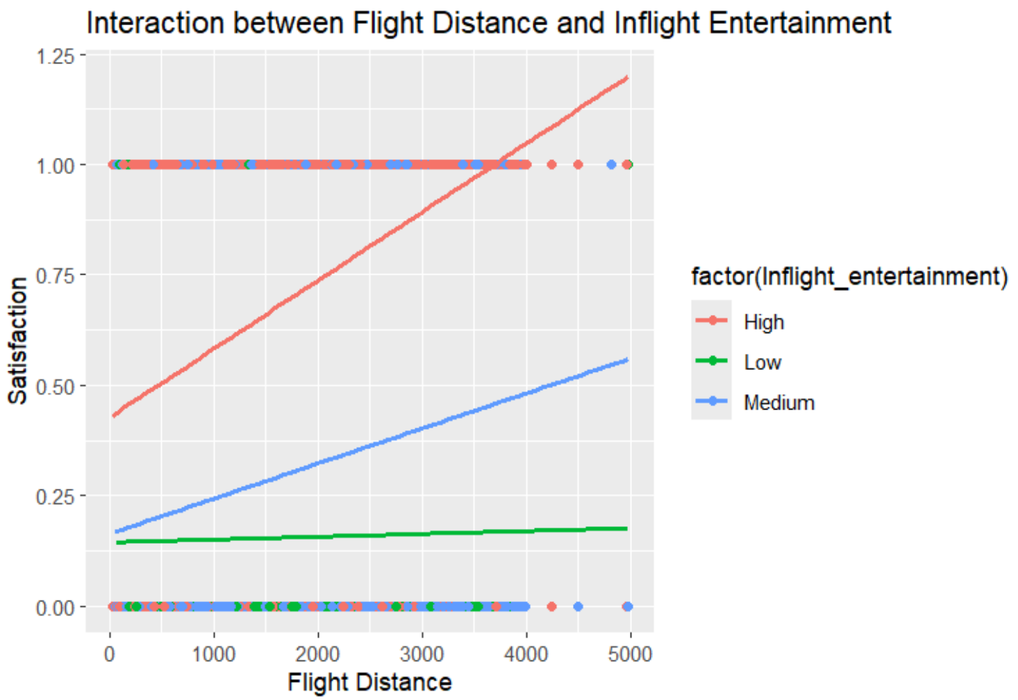
```
> ### 2
> model_dasar2 <- glm(satisfaction ~
+                       Gender +
+                       Age +
+                       Flight_Distance +
+                       Departure_Delay_in_Minutes +
+                       Type_of_Travel +
+                       Class +
+                       Inflight_wifi_service +
+                       Inflight_entertainment,
+                     data = data, family = binomial)
> car::vif(model_dasar2)
                              GVIF Df GVIF^(1/(2*Df))
Gender                    1.002150  1        1.001075
Age                       1.033032  1        1.016382
Flight_Distance           1.211590  1        1.100723
Departure_Delay_in_Minutes 1.004702  1        1.002348
Type_of_Travel            1.273275  1        1.128395
Class                     1.618037  1        1.272021
Inflight_wifi_service     1.315652  2        1.070990
Inflight_entertainment    1.092854  2        1.022446
```

PEMILIHAN MODEL DENGAN METODE STEPWISE FORWARD

```
> # Forward
> model_awal<-glm(satisfaction ~ 1, data = data, family = binomial)
> model_forward <- step(model_awal,
+                       scope = list(lower = model_awal, upper = model_dasar2),
+                       direction = "forward")
Start:  AIC=35624.56
satisfaction ~ 1

                           Df Deviance   AIC
+ Class                     1    28916 28920
+ Type_of_Travel            1    29599 29603
+ Inflight_wifi_service     2    30545 30551
+ Inflight_entertainment    2    30929 30935
+ Flight_Distance           1    33307 33311
+ Age                       1    35236 35240
+ Departure_Delay_in_Minutes 1    35548 35552
<none>                           35623 35625
+ Gender                    1    35621 35625



Step:  AIC=19642.53
satisfaction ~ Class + Inflight_wifi_service + Inflight_entertainment +
    Type_of_Travel + Flight_Distance + Age + Departure_Delay_in_Minutes +
    Gender
```

SUMMARY MODEL FORWARD

```
> summary(model_forward)

Call:
glm(formula = satisfaction ~ Class + Inflight_wifi_service +
    Inflight_entertainment + Type_of_Travel + Flight_Distance +
    Age + Departure_Delay_in_Minutes + Gender, family = binomial,
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9541  -0.5071  -0.1631   0.6121   2.9329

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                    2.397e+00  7.636e-02  31.391  < 2e-16 ***
ClassEco                      -1.699e+00  4.650e-02 -36.545  < 2e-16 ***
Inflight_wifi_serviceLow      -1.505e+00  5.311e-02 -28.332  < 2e-16 ***
Inflight_wifi_serviceMedium   -2.630e+00  4.706e-02 -55.898  < 2e-16 ***
Inflight_entertainmentLow     -2.104e+00  6.639e-02 -31.687  < 2e-16 ***
Inflight_entertainmentMedium  -1.367e+00  3.902e-02 -35.025  < 2e-16 ***
Type_of_TravelPersonal Travel -1.646e+00  5.247e-02 -31.366  < 2e-16 ***
Flight_Distance                2.368e-04  1.938e-05  12.221  < 2e-16 ***
Age                            1.321e-02  1.277e-03  10.344  < 2e-16 ***
Departure_Delay_in_Minutes    -4.007e-03  5.147e-04  -7.785 6.99e-15 ***
GenderMale                     1.232e-01  3.597e-02   3.424 0.000617 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35623  on 25975  degrees of freedom
Residual deviance: 19621  on 25965  degrees of freedom
AIC: 19643

Number of Fisher Scoring iterations: 5
```



INTERAKSI

Grafik ini menunjukkan bahwa kualitas hiburan dapat memperbaiki kepuasan pelanggan, terutama pada penerbangan jarak jauh.

MODEL INTERAKSI

```
> model_2 <- glm(satisfaction ~
+                  Gender +
+                  Age +
+                  Departure_Delay_in_Minutes +
+                  Type_of_Travel +
+                  Class +
+                  Inflight_wifi_service +
+                  Flight_Distance*Inflight_entertainment
+                , data = data, family = binomial)
```

```
> summary((model_2))

Call:
glm(formula = satisfaction ~ Gender + Age + Departure_Delay_in_Minutes +
    Type_of_Travel + Class + Inflight_wifi_service + Flight_Distance *
    Inflight_entertainment, family = binomial, data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2415  -0.5294  -0.1714   0.5722   2.9834

Coefficients:
                                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                                     2.015e+00  7.913e-02  25.465  < 2e-16 ***
GenderMale                                      1.258e-01  3.636e-02   3.460  0.00054 ***
Age                                             1.360e-02  1.283e-03  10.601  < 2e-16 ***
Departure_Delay_in_Minutes                     -4.005e-03  5.238e-04  -7.645 2.08e-14 ***
Type_of_TravelPersonal Travel                  -1.671e+00  5.205e-02 -32.108  < 2e-16 ***
ClassEco                                       -1.678e+00  4.594e-02 -36.518  < 2e-16 ***
Inflight_wifi_serviceLow                       -1.489e+00  5.316e-02 -28.003  < 2e-16 ***
Inflight_wifi_serviceMedium                    -2.673e+00  4.742e-02 -56.369  < 2e-16 ***
Flight_Distance                                 5.695e-04  2.905e-05  19.603  < 2e-16 ***
Inflight_entertainmentLow                      -1.009e+00  9.618e-02 -10.490  < 2e-16 ***
Inflight_entertainmentMedium                   -6.686e-01  6.169e-02 -10.838  < 2e-16 ***
Flight_Distance:Inflight_entertainmentLow      -9.783e-04  6.952e-05 -14.072  < 2e-16 ***
Flight_Distance:Inflight_entertainmentMedium   -5.732e-04  4.001e-05 -14.327  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35623  on 25975  degrees of freedom
Residual deviance: 19278  on 25963  degrees of freedom
AIC: 19304

Number of Fisher Scoring iterations: 5
```

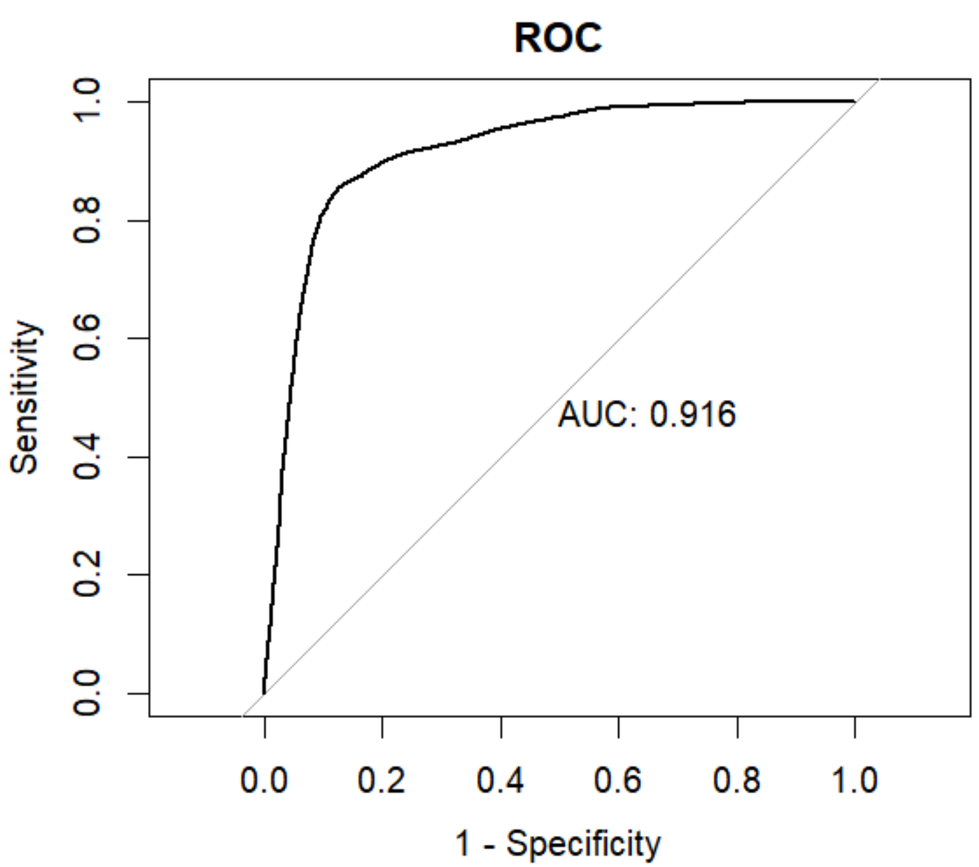
LIKELIHOOD RATIO TEST

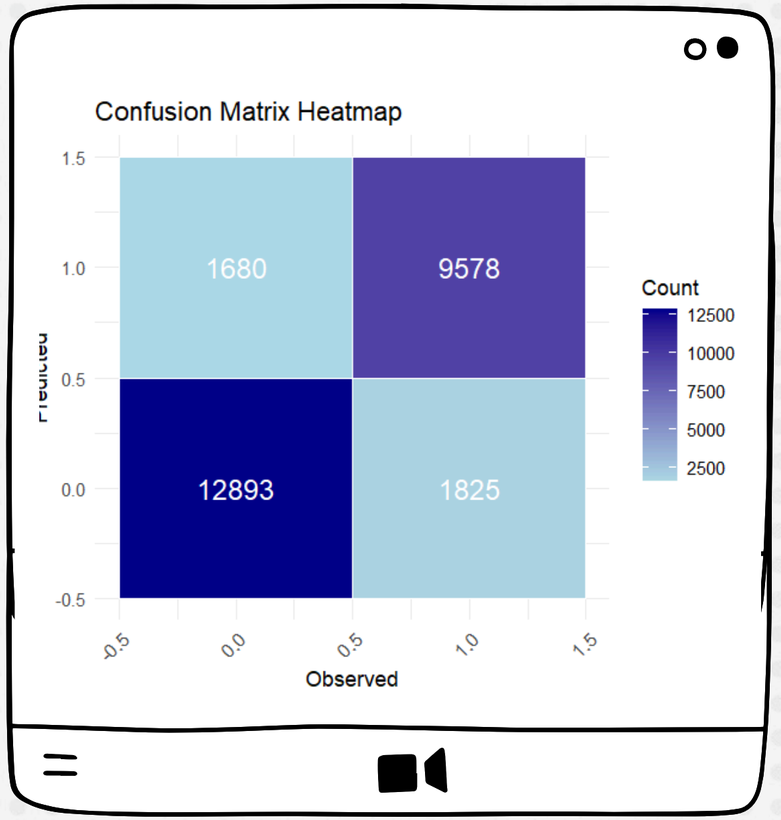
```
> # Membandingkan model interaksi
> anova(model_1,model_2,test = "Chisq")
Analysis of Deviance Table

Model 1: satisfaction ~ Gender + Age + Flight_Distance + Departure_Delay_in_Minutes +
    Type_of_Travel + Class + Inflight_wifi_service + Inflight_entertainment
Model 2: satisfaction ~ Gender + Age + Departure_Delay_in_Minutes + Type_of_Travel +
    Class + Inflight_wifi_service + Flight_Distance * Inflight_entertainment
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     25965      19621
2     25963      19278  2   342.48 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

H0 : Model yang lebih sederhana
H1 : Model yang lebih kompleks

karena p-value < 0.05, maka Tolak H0, artinya model yang lebih kompleks (model_2) dapat menjelaskan data dengan lebih baik, dan menambahkan interaksi memberikan peningkatan yang signifikan.

EVALUASI MODEL





86,5%

AKURASI

MODEL AKHIR

$$
\begin{aligned}
\text{logit}(p) = {} & 2.015 + 1.258 \cdot \textbf{GenderMale} + 1.360 \cdot \textbf{Age} \\
& - 4.005 \cdot \textbf{Departure\_Delay\_in\_Minutes} \\
& - 1.671 \cdot \textbf{Type\_of\_Travel\_Personal\_Travel} \\
& - 1.678 \cdot \textbf{Class\_ECO} \\
& - 1.489 \cdot \textbf{Inflight\_wifi\_service\_Low} \\
& - 0.673 \cdot \textbf{Inflight\_wifi\_service\_Medium} \\
& + 1.695 \cdot \textbf{Flight\_Distance} \\
& - 2.004 \cdot \textbf{Inflight\_entertainment\_Low} \\
& - 1.696 \cdot \textbf{Inflight\_entertainmentMedium} \\
& - 0.0009783 \cdot (\textbf{Flight\_Distance} \cdot \textbf{Inflight\_entertainmentLow}) \\
& - 0.0005732 \cdot (\textbf{Flight\_Distance} \cdot \textbf{Inflight\_entertainmentMedium})
\end{aligned}
$$

$$
\begin{aligned}
\pi &= \frac{e^{logit(\pi)}}{1 + e^{logit(\pi)}} \\
&= \frac{1}{1 + e^{-logit(\pi)}}
\end{aligned}
$$

INTERPRETASI MODEL AKHIR

Dengan interpretasi model sebagai berikut

- Intercept (2.015): Odds dasar kepuasan (ketika semua prediktor = 0) adalah $e$xp(2.015) = 7.50
- Age (0.0136): Setiap kenaikan 1 tahun usia meningkatkan odds kepuasan sebesar exp(0.0136) ≈ 1.014
- Departure_Delay_in_Minutes (-0.004): Setiap penambahan 1 menit keterlambatan menurunkan odds kepuasan sebesar exp(−0.004) ≈ 0.996.
- GenderMale (1.258): Odds kepuasan untuk pria adalah exp(1.258) ≈ 3.52 kalinya wanita, dengan asumsi semua prediktor lain konstan.
- Type_of_Travel_Personal Travel (-1.671): Penumpang yang bepergian untuk keperluan pribadi memiliki odds kepuasan sebesar $e$xp(-1.671) ≈ 0.19 kali lipat dibanding perjalanan bisnis (80% lebih rendah)
- Class(Economy) (-6.678): Penumpang kelas ekonomi memiliki odds kepuasan sebesar exp(-6.678) ≈ 0.0013 kali lipat dibandingkan kelas bisnis.
- Inflight_wifi_service Low (-1.489): Penumpang dengan layanan Wi-Fi rendah memiliki odds kepuasan sebesar exp(-1.489) ≈ 0.23 kali lipat dibandingkan layanan Wi-Fi tinggi.
- Inflight_wifi_service Medium (0.667): Penumpang dengan layanan Wi-Fi Medium memiliki odds kepuasan sebesar exp(0.667) ≈ 1.95 kali lipat dibandingkan layanan Wi-Fi tinggi.

INTERPRETASI MODEL AKHIR

Karena terdapat interaksi antara Flight Distance dengan Inflight Entertaintment:

Jika Inflight_entertainment adalah Low:

- Intercept: 1.006.
  Odds dasar kepuasan adalah $e$xp(1.006) = 2.735
- Flight_Distance: − 0.0004088.
  Setiap kenaikan 1 km jarak penerbangan menurunkan odds kepuasan sebesar exp(− 0.0004088) ≈ 0.9996

Jika Inflight_entertainment adalah Medium

- Intercept: 1.3464
  Odds dasar kepuasan adalah $e$xp(1.3464) = 3.8435
- Flight_Distance: − 0.0000037
  Setiap kenaikan 1 km jarak penerbangan menurunkan odds kepuasan sebesar exp(− 0.0000037) ≈ 0.9999963

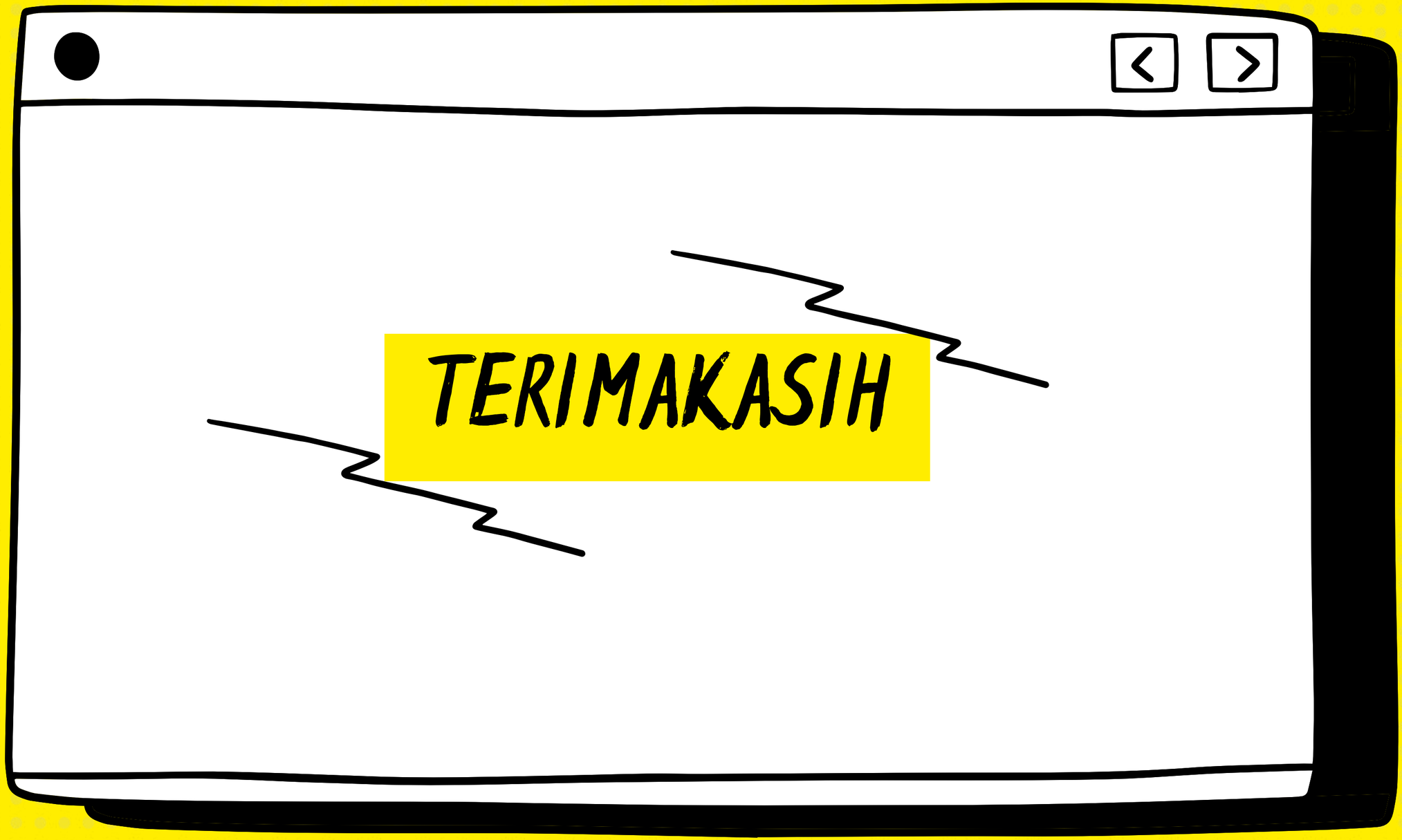
KESIMPULAN

Faktor Peningkat Odds Kepuasan:

- Jenis kelamin pria (odds meningkat 3,52 kali).
- Usia yang lebih tua (kenaikan odds 1,4% per tahun).
- Layanan Wi-Fi menengah (odds meningkat 1,95 kali dibanding layanan rendah).

Faktor Penurun Odds Kepuasan:

- Keterlambatan keberangkatan (penurunan odds kecil tetapi signifikan).
- Perjalanan pribadi (odds hanya 0,19 kali dibanding perjalanan bisnis).
- Kelas ekonomi (odds hanya 0,0013 kali dibanding kelas lainnya).
- Layanan Wi-Fi rendah (odds hanya 0,23 kali dibanding layanan tinggi).


TERIMAKASIH