

ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEPUASAN PENUMPANG PESAWAT MENGGUNAKAN MODEL REGRESI LOGISTIK

Makalah ini disusun untuk memenuhi tugas mata kuliah Analisis Data Kategorik tentang pengaplikasian regresi logistik.

Dosen Pengampu: Madona Yunita Wijaya, M.Sc.



Disusun oleh:
Kelompok 6 DS1

Dani Hidayat	NIM: 11220940000014
Muhammad Fahmi Islam Fadlillah	NIM: 11220940000016
Muhammad Abdul Ghofur	NIM: 11220940000020
Muhammad Anthar Jailani Putra	NIM: 11220940000029
Fadhilah Awal Rizqiansyah	NIM: 11220940000064

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA
2024 M/1445 H

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillah, puji syukur kita panjatkan kepada Allah SWT karena berkat rahmat dan hidayah-Nya kita dapat menyelesaikan Laporan Projek Akhir untuk Mata Kuliah Analisis Data Kategorik yang berjudul “Analisis Faktor-Faktor Yang Mempengaruhi Kepuasan Penumpang Pesawat Menggunakan Model Regresi Logistik”. Shalawat serta salam tidak lupa kita panjatkan kepada junjungan nabi besar Nabi Muhammad SAW. beserta keluarganya, para sahabat, dan para pengikutnya.

Laporan Projek Akhir ini merupakan salah satu kewajiban yang harus dilaksanakan oleh mahasiswa yang mengikuti Mata Kuliah Analisis Data Kategorik Program Studi Matematika Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta sebagai pengimplementasian teori yang sudah dipelajari selama perkuliahan berlangsung dan juga sebagai penilaian Ujian Akhir Semester.

Kami menyadari bahwa dalam penulisan makalah ini tidak terlepas dari bantuan banyak pihak, oleh karena itu kami mengucapkan terima kasih dan mohon maaf apabila ada kekurangan dalam penulisan laporan ini.

Wassalamualaikum Warahmatullahi Wabarakatuh

Tangerang Selatan, 20 Desember 2024

Penulis

ABSTRAK

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi tingkat kepuasan penumpang pesawat berdasarkan survei data kepuasan pelanggan maskapai penerbangan. Data yang digunakan meliputi variabel seperti jenis kelamin, usia, jenis perjalanan, kelas penerbangan, jarak penerbangan, keterlambatan keberangkatan, keterlambatan kedatangan, layanan WiFi, dan hiburan selama penerbangan.

Metode analisis menggunakan model regresi logistik untuk mengidentifikasi faktor signifikan yang mempengaruhi kepuasan. Hasil penelitian menunjukkan bahwa jenis kelamin pria, usia penumpang yang lebih tua, dan layanan WiFi berkualitas medium berpengaruh positif terhadap kepuasan penumpang. Sebaliknya, keterlambatan penerbangan, perjalanan pribadi, kelas ekonomi, layanan WiFi rendah, dan kualitas hiburan yang buruk secara signifikan menurunkan tingkat kepuasan penumpang. Selain itu, ditemukan interaksi antara jarak penerbangan dan kualitas hiburan, yang menekankan pentingnya hiburan pada penerbangan jarak jauh.

Dengan akurasi model sebesar 86,5%, penelitian ini memberikan wawasan bagi maskapai untuk meningkatkan kepuasan pelanggan melalui perbaikan layanan yang lebih fokus pada aspek signifikan seperti peningkatan kualitas layanan hiburan, WiFi, dan pengelolaan ketepatan waktu keberangkatan.

Kata Kunci: Kepuasan penumpang, regresi logistik, maskapai penerbangan, layanan penerbangan, kualitas hiburan

DAFTAR ISI

KATA PENGANTAR.....	1
ABSTRAK.....	2
DAFTAR ISI.....	3
BAB 1 PENDAHULUAN.....	4
1.1 LATAR BELAKANG	4
1.2 RUMUSAN MASALAH	4
1.3 TUJUAN DAN MANFAAT PENULISAN.....	4
BAB 2.....	5
LANDASAN TEORI	5
2.1 KEPUASAN MASKAPAI PENERBANGAN	5
2.2 REGRESI LOGISTIK	5
2.3 MODEL PERSAMAAN REGRESI LOGISTIK	6
2.4 ELIMINASI BACKWARD	7
2.5 SELEKSI FORWARD	7
2.6 UJI KELAYAKAN MODEL (<i>GOODNESS OF FIT</i>)	7
2.7 INFORMATION CRITERIAN (AIC).....	8
2.8 TABEL KLASIFIKASI	8
2.9 KURVA ROC.....	8
BAB 3.....	10
METODE PENELITIAN	10
3.1 SUMBER DATA.....	10
3.1 SUMBER DATA	10
3.2 VARIABEL PENELITIAN.....	10
3.3 ANALISIS DATA	11
BAB 4.....	12
HASIL DAN PEMBAHASAN	12
4.1 MODEL REGRESI BERGANDA	12
4.2 UJI FIT MODEL	14
4.2.1 Uji Klasifikasi	14
4.2.2 Kurva ROC.....	15
4.2.3 Likelihood Ratio Test	15
4.3 MODEL AKHIR.....	16
BAB 5.....	17
KESIMPULAN.....	17
DAFTAR PUSTAKA.....	19
APPENDIKS.....	20

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dalam era perkembangan maskapai penerbangan yang semakin kompetitif, memahami faktor-faktor yang mempengaruhi kepuasan penumpang menjadi suatu hal yang sangat penting untuk dapat meningkatkan layanan.

Dataset yang digunakan dalam analisis ini merupakan hasil survei kepuasan pelanggan maskapai penerbangan, yang berasal dari sumber internet. Dengan variabel targetnya yaitu tingkat kepuasan pelanggan yang terbagi menjadi puas dan tidak puas.

Analisis regresi logistik biner merupakan metode statistik yang digunakan untuk memprediksi variabel respons biner (dalam kasus ini “puas” atau “tidak puas”) berdasarkan variabel prediktor yang bersifat kategorik maupun numerik. Oleh karena itu, digunakan analisis regresi logistik biner untuk memprediksi dan mengklasifikasikan kepuasan penumpang maskapai penerbangan.

Hasil dari analisis ini dapat memberikan wawasan berharga bagi maskapai penerbangan dalam meningkatkan layanan mereka serta memenuhi preferensi pelanggan, sehingga memperkuat posisi mereka dalam industri penerbangan yang kompetitif.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, permasalahan yang akan dibahas yaitu analisis data untuk menemukan faktor-faktor yang berhubungan dengan tingkat kepuasan penumpang maskapai penerbangan.

1.3 Tujuan dan Manfaat Penulisan

Tujuan dari penelitian ini adalah untuk meneliti dan mengidentifikasi faktor-faktor yang memiliki pengaruh signifikan terhadap tingkat kepuasan penumpang dalam maskapai penerbangan yang kompetitif. Sedangkan manfaat dari penelitian ini adalah untuk mengetahui dan memahami penerapan regresi logistik dalam memberikan wawasan yang mendalam kepada maskapai penerbangan untuk meningkatkan kualitas layanan mereka, memberikan pengalaman pelanggan yang lebih memuaskan, dan memperbaiki kelemahan yang mungkin a

BAB 2

LANDASAN TEORI

2.1 Kepuasan Maskapai Penerbangan

Kepuasan penumpang dalam industri penerbangan dapat dipahami melalui sejumlah faktor kunci yang mencerminkan pengalaman perjalanan secara keseluruhan. Pertama-tama, jenis pesawat dan kelas penerbangan, seperti ekonomi atau ekonomi plus, memainkan peran signifikan dalam membentuk persepsi penumpang. Kualitas kursi dan fasilitas tambahan menjadi aspek penting yang memengaruhi kepuasan mereka.

Selanjutnya, faktor waktu juga memiliki dampak yang besar. Pengalaman penerbangan yang nyaman terkait dengan waktu keberangkatan dan kedatangan yang tepat dapat meningkatkan kepuasan penumpang. Efisiensi dalam proses boarding dan disembarkasi juga memberikan kontribusi positif terhadap pengalaman perjalanan.

Pelayanan dalam pesawat mencakup layanan Wifi dan layanan hiburan adalah variabel lain yang perlu diperhatikan. Ketersediaan WiFi dan layanan hiburan (termasuk film, musik, atau konten lainnya) menciptakan suasana perjalanan yang lebih menyenangkan dan dapat meningkatkan kualitas pengalaman perjalanan penumpang.

Terakhir, jarak penerbangan juga berperan penting dalam menentukan kepuasan penumpang. Untuk penerbangan dengan jarak tempuh yang jauh, maskapai perlu memastikan bahwa fasilitas di pesawat dapat memenuhi kebutuhan penumpang dalam waktu yang lama.

Secara keseluruhan, pemahaman yang mendalam terhadap variabel-variabel ini membantu maskapai penerbangan untuk merancang pengalaman perjalanan yang lebih baik. Dengan memperhatikan kelas penerbangan, waktu perjalanan, layanan Wifi, hiburan di pesawat, dan jarak penerbangan, maskapai dapat meningkatkan kualitas layanan mereka dan memenuhi harapan penumpang, yang pada akhirnya dapat meningkatkan tingkat kepuasan penumpang secara keseluruhan.

2.2 Regresi Logistik

Regresi logistik adalah sebuah pendekatan untuk membuat model prediksi seperti halnya regresi linear atau yang biasa disebut dengan istilah Ordinary Least Squares (OLS) regression. Perbedaannya adalah pada regresi logistik, peneliti memprediksi variabel terikat yang berskala dikotomi. Skala dikotomi yang dimaksud adalah skala data nominal dengan dua kategori, misalnya ya dan tidak, baik dan buruk atau tinggi dan rendah. Seperti semua analisis regresi, regresi logistik adalah analisis prediktif. Regresi ini digunakan untuk menggambarkan data dan untuk menjelaskan hubungan antara satu variabel binary dependen dan variabel independen nominal, ordinal, interval, atau rasio.

Tetapi tidak hanya itu regresi logistik juga biasa digunakan untuk menghitung terjadinya peristiwa binary dan untuk menangani masalah klasifikasi. Misalnya memprediksi apakah email yang masuk adalah spam atau bukan. Dalam konteks medis atau kesehatan, regresi ini dapat digunakan untuk memprediksi apakah penyakit tersebut berbahaya atau tidak, contohnya yaitu memprediksi tumor itu jinak atau ganas. Intinya, regresi ini biasa digunakan untuk memprediksi kemungkinan semua jenis hasil

“ya” atau “tidak”. Dengan memprediksi hasil seperti itu, ini membantu analisis data untuk membuat keputusan yang tepat. Dan mungkin dalam skema besar suatu bisnis ini sangat membantu meminimalkan risiko kerugian dan mengoptimalkan pengeluaran untuk memaksimalkan keuntungan.

Ada beberapa persyaratan atau asumsi yang harus dipenuhi suatu data agar regresi logistik bisa diaplikasikan dengan benar. Asumsi-asumsi yang dimaksud diantaranya:

- Regresi logistik tidak membutuhkan hubungan linier antara variabel independen dengan variabel dependen Untuk menggunakan analisis regresi ini, jenis data yang dimiliki oleh peneliti harus bersifat dikotomi, yakni hanya ada dua kategori saja.
- Variabel independen tidak memerlukan asumsi multivariate normality.
- Asumsi homokedastisitas tidak diperlukan.
- Variabel bebas tidak perlu diubah ke dalam bentuk metrik (interval atau skala ratio).
- Variabel dependen harus bersifat dikotomi (2 kategori, misal tinggi dan rendah atau baik dan buruk).
- Variabel independen tidak harus memiliki keragaman yang sama antar kelompok variabel.
- Kategori dalam variabel independen harus terpisah satu sama lain atau bersifat eksklusif.
- Sampel yang diperlukan dalam jumlah relatif besar, minimum dibutuhkan hingga 50 sampel data untuk sebuah variabel prediktor (independen).
- Dapat menyeleksi hubungan karena menggunakan pendekatan non linier log transformasi untuk memprediksi odds ratio. Odd dalam regresi logistik sering dinyatakan sebagai probabilitas.

2.3 Model Persamaan Regresi Logistik

Model persamaan aljabar layaknya OLS yang biasa digunakan adalah

$$y = \beta_0 + \beta_1 X + \varepsilon$$

dimana ε adalah error varians atau residual. Dengan model regresi ini, tidak menggunakan interpretasi yang sama seperti halnya persamaan regresi OLS. Model persamaan yang terbentuk berbeda dengan persamaan OLS. Persamaannya yaitu

$$\ln \left(\frac{\hat{p}}{1 - \hat{p}} \right) = \beta_0 + \beta_1 X$$

- \ln adalah logaritma natural
- $\beta_0 + \beta_1 X$ adalah persamaan yang biasa dikenal dalam OLS
- p adalah probabilitas logistik yang didapat dari rumus

$$\hat{p} = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

dimana \exp atau ditulis “e” adalah fungsi eksponen. Perlu diingat bahwa eksponen merupakan kebalikan dari logaritma natural. Sedangkan logaritma natural adalah bentuk logaritma namun dengan nilai konstanta 2,71828182845904 atau bisa dibulatkan menjadi 2,72.

Regresi logistik itu sendiri terbagi menjadi 2 yaitu regresi logistik sederhana dan regresi logistik berganda. Model persamaan regresi logistik sederhana adalah

$$\text{logit}(\pi) = \log \left(\frac{\pi}{1 - \pi} \right) = a + \beta x$$

Fungsi logit memetakan bilangan diantara 0 dan 1 ke bilangan diantara $-\infty$ dan $+\infty$. Adapun peluang sukses dapat dihitung sebagai

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

Lalu model persamaan untuk regresi logistik berganda adalah

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = a + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Fungsi logit memetakan bilangan diantara 0 dan 1 ke bilangan diantara $-\infty$ dan $+\infty$. Adapun peluang sukses dapat dihitung sebagai

$$\pi(x) = \frac{e^{a + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{a + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}$$

2.4 Eliminasi Backward

Metode Backward Elimination dilakukan dengan cara memasukkan semua predictor kemudian mengeliminasi satu persatu hingga tersisa prediktor yang signifikan saja. Eliminasi ini didasarkan pada prediktor yang memiliki nilai sig F yang di atas 0.1. Berikut ini adalah langkah-langkah melakukan Eliminasi Backward:

1. Mulai dengan model yang kompleks (meliputi semua variabel independent dan interaksi diantara variabel independent).
2. Buang variabel “yang paling tidak signifikan” (p-value terbesar).
 - Jangan buang pengaruh utama tanpa membuang pengaruh interaksi yang berkaitan.
 - Jangan buang hanya salah satu variabel dummy yang berasal dari variable independent kategorik.
3. Fit model kembali.
4. Diulangi sampai diperoleh semua variabel “signifikan”.

2.5 Seleksi Forward

Metode forward selection dilakukan dengan cara memasukkan prediktor secara bertahap, prediktor ini berdasarkan korelasi parsial terbesar. Dalam metode forward selection, variabel prediktor yang dimasukkan dalam model tidak akan dapat dikeluarkan lagi. Proses tersebut dihentikan ketika prediktor-prediktor baru tidak bisa meningkatkan berpengaruh secara signifikan (sig di bawah 0.05) terhadap variable respon. Karena itulah prosedur forward selection menjadi salah satu prosedur pemilihan model terbaik dalam regresi dengan eliminasi variabel bebas yang membangun model secara bertahap.

Ada beberapa cara yang dapat digunakan dalam pengujian dengan metode forward selection ini. Berikut langkah-langkah melakukan seleksi forward:

- Mulai dari model hanya dengan intercept (~1).
- Variabel “yang paling signifikan” dimasukan ke dalam model di setiap langkah sampai tidak ada lagi variabel yang signifikan. Untuk melakukan seleksi forward, harus memberikan argument “scope”.

2.6 Uji Kelayakan Model (Goodness Of Fit)

Uji kelayakan model bertujuan untuk mengetahui ketepatan fungsi regresi sampel dalam menaksir nilai aktual. Untuk menguji kelayakan model regresi dinilai dengan menggunakan uji Hosmer-Lemeshow’s Goodness of Fit yang diukur dengan nilai Chi Square. Hosmer-Lemeshow’s Goodness of Fit menguji hipotesis nol bahwa data tidak ada perbedaan signifikan antara model dengan data sehingga model dapat dikatakan fit. (Ghozali, 2018:331)

Statistik Hosmer-Lemeshow's dihitung sebagai Pearson chi-squared dari table kontingensi frekuensi aktual dan frekuensi harapan. Seperti halnya asosiasi table 2 arah, model dikatakan fit jika uji Hosmer-Lemeshow's memberikan nilai p-value yang semakin besar. Ketika terdapat variabel kontinu, maka akan terdapat banyak sel yang didefinisikan oleh variabel independent, membuat tabel kontingensi semakin besar, sehingga akan lebih sering memberikan kemungkinan hasil yang signifikan.

Jika uji Hosmer-Lemeshow's menunjukkan nilai probabilitas ($P - value$) ≤ 0.05 artinya adanya perbedaan signifikan antara model dengan nilai observasinya sehingga model tidak dapat digunakan untuk memprediksi nilai observasinya.

Jika uji Hosmer-Lemeshow's menunjukkan nilai probabilitas ($P - value$) ≥ 0.05 artinya tidak adanya perbedaan signifikan antara model dengan data atau bisa dikatakan model dapat digunakan untuk memprediksi nilai observasinya.

2.7 Information Criterion (AIC)

AIC adalah kriteria pemilihan model yang memilih model yang meminimumkan :

$$AIC = -2((\max. \loglikelihood) + 2(banyaknya parameter))$$

AIC cenderung memilih model sederhana (sedikit parameter) sebagai model yang fit. Bisa digunakan untuk membandingkan model yang bukan merupakan subset dari suatu model lainnya. Menurut metode AIC, model regresi yang baik adalah model regresi Ketika nilai AIC-nya kecil.

2.8 Tabel Klasifikasi

Tabel klasifikasi merupakan tabel yang memuat pengelompokan data, dapat berupa tabel klasifikasi tunggal dan ganda. Tabel klasifikasi digunakan untuk mengetahui seberapa besar klasifikasi yang dihasilkan dengan cara membandingkan nilai aktual dengan nilai prediksi. Dalam tabel klasifikasi terdapat istilah sensitivity dan specificity. Sensitivity adalah proporsi data pengamatan positif yang diklasifikasikan positif oleh model. Specificity adalah proporsi data pengamatan negatif yang diklasifikasikan negatif oleh model.

2.9 Kurva ROC

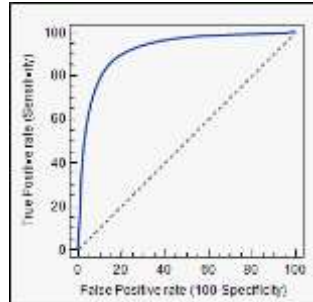
Kurva ROC adalah representasi grafis dari hubungan antara sensitivitas dan 1-spesifisitas. Dalam penelitian medis kurva ROC banyak digunakan untuk menggambarkan keakuratan diagnostik dan menentukan nilai cut-off yang optimal. Keakuratan diagnosis berasal dari area di bawah kurva ROC dan optimal cut-off digunakan untuk mengidentifikasi kondisi positif dan negatif dalam diagnosis. Banyak penelitian telah menggunakan kurva ROC dengan metode empiris untuk menggambarkan akurasinya.

Kurva ROC empiris mempertahankan sifat-sifat dari fungsi distribusi empiris yang tidak bias terhadap distribusi teoritisnya, tetapi tidak sangat akurat ketika ukuran sampel kecil karena variabilitas dari luasan bawah kurva ROC. Metode lain diperlukan untuk mendapatkan kurva ROC yang cukup akurat untuk ukuran sampel yang kecil. Kemudian, kurva ROC smoothed empirical diusulkan untuk menangani akurasi dari kurva ROC dalam menggambarkan akurasi diagnostik, terutama untuk ukuran sampel yang kecil.

Selanjutnya, dengan estimator smoothed empirical dapat memperoleh nilai sensitivitas dan 1-spesifisitas dalam membangun kurva untuk membentuk daerah di

bawah kurva ROC menggunakan metode rotate ordinal graph. Setelah itu, area di bawah kurva ROC dapat diukur dengan menggunakan aturan trapesium.

- Kurva ROC meringkas performa model dengan mengevaluasi trade-off antara nilai sensitivity dan specificity. Sensitivity adalah proporsi data pengamatan positif yang diklasifikasikan positif oleh model, Sedangkan Specificity adalah proporsi data pengamatan negatif yang diklasifikasikan negatif oleh model.



Gambar 2.9

- Salah satu cara menentukan aturan klasifikasi adalah menentukan titik potong c ; kemudian mengklasifikasikan respon data dengan prediksi peluang diatas c sebagai positif dan kurang atau sama dengan c sebagai negatif.
- Jika titik potong c semakin besar maka sedikit respon ($Y = 1$) yang diprediksi positif (sensitivity semakin berkurang), tapi banyak respon ($Y = 0$) yang diprediksi negatif (specificity semakin bertambah).
- Model dengan kemampuan diskriminasi yang tinggi akan memiliki sensitivity dan specificity tinggi secara bersamaan. Kemampuan diskriminasi tinggi ditunjukkan oleh kurva yang semakin mendekati ujung kiri atas plot, sedangkan model tidak memiliki kemampuan diskriminasi jika kurva ROC dekat dengan garis dengan kemiringan 45 derajat.
- Area di bawah kurva ROC disebut AUC (Area Under the Curve) adalah ukuran untuk meringkas kemampuan diskriminasi model. Nilai AUC berada pada rentang 1 (diskriminasi sempurna) dan 0.5 (tidak ada kemampuan diskriminasi).

BAB 3 METODE PENELITIAN

Mengacu pada bentuk penelitian, tujuan, sifat masalah dan pendekatannya, maka kami menggunakan metode penelitian deskriptif. Metode penelitian deskriptif adalah metode yang digunakan untuk mencari unsur-unsur, ciri-ciri, sifat-sifat suatu fenomena. Metode ini dimulai dengan mengumpulkan data, menganalisis data, dan menginterpretasikannya. Sumber data yang diperoleh adalah data sekunder yang bersumber dari Kaggle yang dapat diakses di <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

3.2 Variabel Penelitian

Variabel yang akan diamati dalam penelitian ini terdiri dari 1 variabel dependent dan 13 variabel independent yaitu:

- a. Variabel Y (variabel dependent): **Satisfaction** (Tingkat kepuasan terhadap maskapai penerbangan). Karena fokus penelitian ini adalah binary maka variabel respon harus terdiri dari hanya 2 kategori. Sehingga diperoleh data dari variabel tersebut yaitu:
Y = 0, penumpang memberikan nilai **neutral or dissatisfied**
Y = 1, penumpang memberikan nilai **satisfied**
- b. Variabel bebas (variabel independent) yang digunakan terdiri dari 9 variabel yaitu:
 1. Age (X₁) merupakan variabel numerik yang memuat usia sebenarnya dari penumpang.
 2. Flight Distance (X₂) merupakan variabel numerik yang memuat jarak penerbangan perjalanan.
 3. Departure Delay In Minutes (X₃) merupakan variabel numerik yang memuat lama keterlambatan keberangkatan penerbangan dalam menit.
 4. Arrival Delay In Minutes (X₄) merupakan variabel numerik yang memuat lama keterlambatan kedatangan penerbangan dalam menit.
 5. Gender (X₅) merupakan variabel kategorik yang memuat jenis kelamin penumpang penerbangan (Wanita, Pria).
 6. Type Of Travel (X₆) merupakan variabel kategorik yang memuat tujuan perjalanan penumpang (Personal Travel: Perjalanan Pribadi, Business Travel: Perjalanan Bisnis).
 7. Class (X₇) merupakan variabel kategorik yang memuat tipe kelas perjalanan di pesawat (Business: Bisnis, Eco: Ekonomi, Eco Plus: Ekonomi Plus)
 8. Inflight Wifi Service (X₈) merupakan variabel kategorik yang memuat tingkat kepuasan terhadap layanan WiFi di dalam penerbangan (0: Tidak Berlaku; 1-5: Skala Kepuasan).
 9. Inflight entertainment (X₉) merupakan variabel kategorik yang memuat tingkat kepuasan terhadap hiburan di dalam penerbangan (1-5: Skala Kepuasan).

3.3 Analisis Data

Langkah-langkah yang dilakukan dalam proses penelitian adalah sebagai berikut

1. Pengambilan data.
2. Melakukan analisis regresi logistik.
3. Melakukan pemilihan model terbaik.
4. Menginterpretasikan model.
5. Membuat kesimpulan dan saran hasil penelitian.

BAB 4 HASIL DAN PEMBAHASAN

Sebelum masuk ke pemodelan, pertama akan terlebih dahulu dilakukan data preprocessing menggunakan Python, dimana dalam tahap ini dilakukan penanganan terhadap missing value pada data, membuang variabel-variabel yang tidak dibutuhkan, dan mengganti tipe beberapa variabel. Masih menggunakan Python, selanjutnya ditampilkan visualisasi yang memuat informasi menarik dari beberapa variabel. Hasil dari visualisasi terlampir pada analisis deskriptif sebagai berikut

4.1 Model Regresi Berganda

Dalam analisis ini, variabel **Satisfaction** digunakan sebagai variabel dependen (Y), sedangkan variabel lainnya berperan sebagai variabel independen. Analisis dilakukan menggunakan metode regresi logistik berganda untuk mengevaluasi pengaruh variabel independen terhadap tingkat kepuasan. Setelah model awal diperoleh, dilakukan prosedur **Stepwise Forward Elimination** dengan bantuan fungsi stepAIC di R untuk menyaring variabel-variabel yang signifikan dan mendapatkan model yang lebih optimal dan sesuai dengan data. Proses ini bertujuan untuk menghasilkan model yang parsimonious, yaitu model yang sederhana namun tetap mampu menjelaskan variasi dalam data secara memadai. Adapun model yang diperoleh sebagai berikut.

```
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
```

Dari output di atas diperoleh persamaan regresi logistik yaitu:

$$\begin{aligned} \text{satisfaction} = & 2.397 - 1.699(\text{ClassEco}) \\ & - 1.505(\text{Inflight_wifi_serviceLow}) \\ & - 2.630(\text{Inflight_wifi_serviceMedium}) \end{aligned}$$

$$\begin{aligned}
& - 2.104(\text{Inflight_entertainmentLow}) \\
& - 1.367 (\text{Inflight_entertainmentMedium}) \\
& - 1.646(\text{Type_ofTravelPersonal Travel}) \\
& + 0.0002368(\text{Flight_Distance}) \\
& + 0.01321(\text{Age}) \\
& - 0.004007(\text{Departure_Delay_in_Minutes}) \\
& + 0.1232(\text{GenderMale})
\end{aligned}$$

Setelah model awal diperoleh, dilakukan analisis lebih lanjut untuk mengevaluasi hubungan antar variabel independen. Dalam hal ini, kami mengasumsikan bahwa variabel **Flight_Distance** dan **Inflight_Entertainment** memiliki korelasi, di mana perjalanan yang lebih jauh disertai dengan hiburan yang menarik selama penerbangan dapat meningkatkan tingkat kepuasan penumpang secara signifikan. Asumsi ini didasarkan pada pengalaman pengguna yang cenderung merasa lebih nyaman dan puas ketika perjalanan panjang dilengkapi dengan fasilitas hiburan yang berkualitas. Oleh karena itu, dilakukan analisis interaksi antara kedua variabel tersebut untuk menguji pengaruh gabungannya terhadap tingkat kepuasan penumpang. Adapun model interaksi yang diperoleh dari analisis ini adalah sebagai berikut.

```

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

```

Dari output diperoleh model dengan variabel yang lebih signifikan yaitu:

$$\begin{aligned}
\text{satisfaction} = & 2.015 - 1.678(\text{ClassEco}) \\
& + 0.1258(\text{GenderMale}) \\
& + 0.01321(\text{Age}) \\
& - 0.004005(\text{Departure_Delay_in_Minutes}) \\
& - 1.671(\text{Type_ofTravelPersonal Travel}) \\
& - 1.489(\text{Inflight_wifi_serviceLow}) \\
& - 2.673(\text{Inflight_wifi_serviceMedium}) \\
& + 0.0005695(\text{Flight_Distance})
\end{aligned}$$

- 1.009(*Inflight_entertainmentLow*)
- 0.6686(*Inflight_entertainmentMedium*)
- 0.0009783(*Flight_Distance: Inflight_entertainmentLow*)
- 0.0005732(*Flight_Distance: Inflight_entertainmentMedium*)

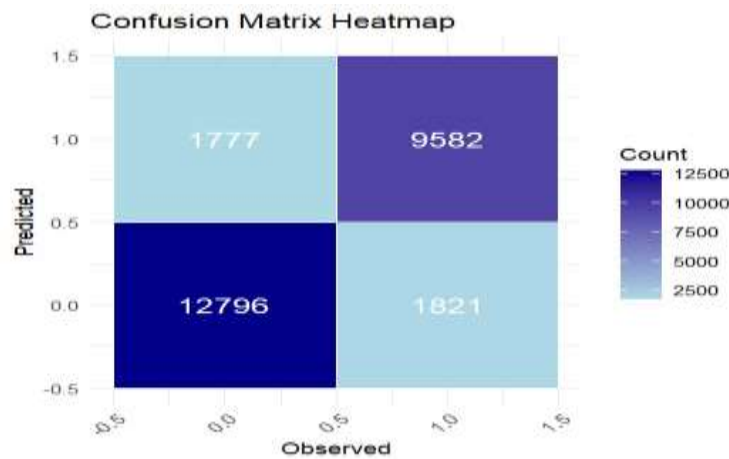
4.2 Uji Fit Model

Setelah selesai dalam pemodelan, langkah selanjutnya adalah memeriksa apakah model tersebut sudah fit atau tidak fit. Banyak cara atau metode untuk memeriksanya namun pada kali ini akan digunakan 2 metode yaitu Tabel Klasifikasi, dan Kurva ROC.

Uji Klasifikasi

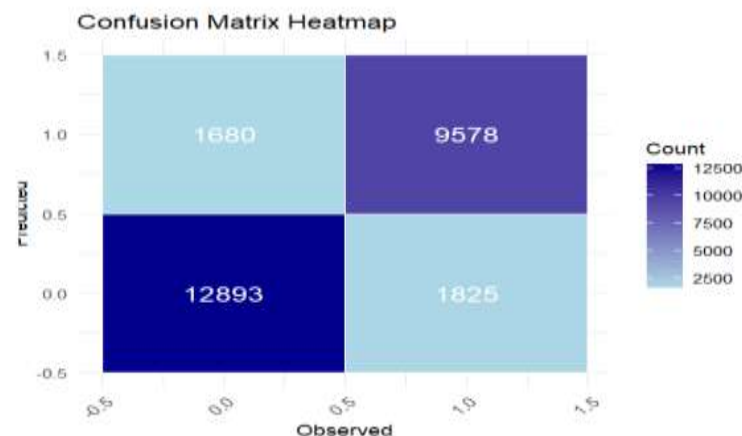
Metode yang digunakan selanjutnya adalah tabel klasifikasi yang bertujuan untuk mendapatkan nilai akurasi.

model₁:



Gambar 4.1a

model₂:



Gambar 4.1b

Dari output di atas dapat dihitung nilai akurasinya sebagai berikut

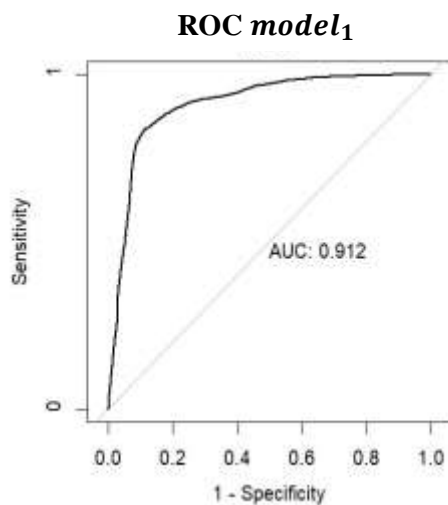
$$\text{akurasi model}_1 = \frac{12796 + 9582}{25976} = 0.8614875$$

$$akurasi\ model_2 = \frac{12893+9578}{25976} = 0.865068$$

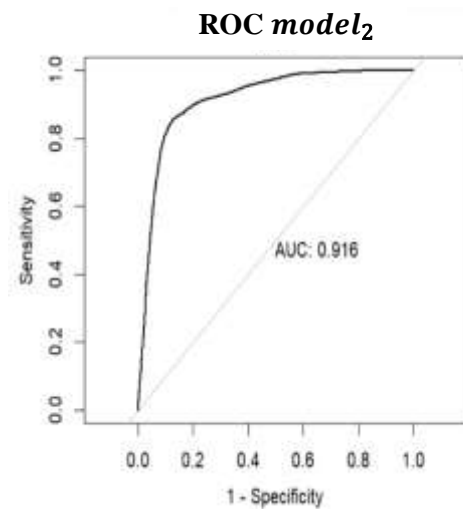
Sehingga diperoleh bahwa akurasi dari **model₂** lebih baik dari pada **model₁** yaitu sebesar 86,1% dengan error sebesar 13,9%.

Kurva ROC

Dalam kurva ROC akan ditunjukkan kemampuan diskriminasi suatu model, jika model memiliki nilai sensitivity dan specificity yang tinggi maka kemampuan diskriminasi model juga tinggi.



Gambar 4.2a



Gambar 4.2b

Dari garfuk kurva ROC dari **model₂** yang menunjukkan performa suatu klasifikasi biner. Pada kurva ini, nilai AUC (*Area Under Curve*) **model₂** sebesar 0.916, yang menunjukkan tingkat kinerja yang lebih baik dari nilai AUC **model₁**.

Nilai AUC yang semakin mendekati 1 menunjukkan kinerja klasifikator yang semakin baik, sedangkan nilai yang mendekati 0.5 menandakan klasifikator yang tidak lebih baik dari tebakan acak.

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

Berdasarkan hasil uji, diperoleh nilai p-value < 0,05. Dengan demikian, **H0** ditolak, yang berarti model yang lebih kompleks **model₂** memberikan penjelasan yang lebih baik terhadap data. Hal ini menunjukkan bahwa penambahan interaksi dalam model memberikan peningkatan yang signifikan dalam kemampuan model untuk menjelaskan variabilitas data.

4.3 Model Akhir

Persamaan regresi logistik yang memiliki signifikan cukup tinggi adalah

$$\begin{aligned} \text{Satisfaction} = & 2.015 - 1.678(\text{ClassEco}) \\ & + 0.1258(\text{GenderMale}) \\ & + 0.01321(\text{Age}) \\ & - 0.004005(\text{Departure_Delay_in_Minutes}) \\ & - 1.671(\text{Type_ofTravelPersonal Travel}) \\ & - 1.489(\text{Inflight_wifi_serviceLow}) \\ & - 2.673(\text{Inflight_wifi_serviceMedium}) \\ & + 0.0005695(\text{Flight_Distance}) \\ & - 1.009(\text{Inflight_entertainmentLow}) \\ & - 0.6686(\text{Inflight_entertainmentMedium}) \\ & - 0.0009783(\text{Flight_Distance: Inflight_entertainmentLow}) \\ & - 0.0005732(\text{Flight_Distance: Inflight_entertainmentMedium}) \end{aligned}$$

Karena terdapat interaksi pada model, maka model dapat dipecah menjadi dua. Model untuk *Inflight entertainment Low* dan *Inflight entertainment Medium*

$$\begin{aligned} \text{Satisfaction_Inflight entertainment Low} = & 1.006 - 1.678(\text{ClassEco}) \\ & + 0.1258(\text{GenderMale}) \\ & + 0.01321(\text{Age}) \\ & - 0.004005(\text{Departure_Delay_in_Minutes}) \\ & - 1.671(\text{Type_ofTravelPersonal Travel}) \\ & - 1.489(\text{Inflight_wifi_serviceLow}) \\ & - 2.673(\text{Inflight_wifi_serviceMedium}) \\ & - 0.0004088(\text{Flight_Distance}) \end{aligned}$$

$$\begin{aligned} \text{Satisfaction_Inflight entertainment Medium} = & 1.3464 \\ & - 1.678(\text{ClassEco}) \\ & + 0.1258(\text{GenderMale}) \\ & + 0.01321(\text{Age}) \\ & - 0.004005(\text{Departure_Delay_in_Minutes}) \\ & - 1.671(\text{Type_ofTravelPersonal Travel}) \\ & - 1.489(\text{Inflight_wifi_serviceLow}) \\ & - 2.673(\text{Inflight_wifi_serviceMedium}) \\ & - 0.0000037(\text{Flight_Distance}) \end{aligned}$$

BAB 5 KESIMPULAN

Berdasarkan proses penelitian yang telah dilakukan mulai dari analisis deskriptif data, analisis regresi logistik hingga evaluasi model dapat disimpulkan sebagai berikut.

1. Dengan interpretasi model sebagai berikut

- *Intercept* (2.015): Odds dasar kepuasan (ketika semua prediktor = 0) adalah $\exp(2.015) = 7.50$
- *Age* (0.0136): Setiap kenaikan 1 tahun usia meningkatkan odds kepuasan sebesar $\exp(0.0136) \approx 1.014$
- *Departure_Delay_in_Minutes* (-0.004): Setiap penambahan 1 menit keterlambatan menurunkan odds kepuasan sebesar $\exp(-0.004) \approx 0.996$.
- *GenderMale* (1.258): Odds kepuasan untuk pria adalah $\exp(1.258) \approx 3.52$ 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 $\exp(-1.671) \approx 0.19$ kali lipat dibanding perjalanan bisnis (80% lebih rendah)
- *Class(Economy)* (-6.678): Penumpang kelas ekonomi memiliki odds kepuasan sebesar $\exp(-6.678) \approx 0.0013$ kali lipat dibandingkan kelas bisnis.
- *Inflight_wifi_service Low* (-1.489): Penumpang dengan layanan Wi-Fi rendah memiliki odds kepuasan sebesar $\exp(-1.489) \approx 0.23$ kali lipat dibandingkan layanan Wi-Fi tinggi.
- *Inflight_wifi_service Medium* (0.667): Penumpang dengan layanan Wi-Fi Medium memiliki odds kepuasan sebesar $\exp(0.667) \approx 1.95$ kali lipat dibandingkan layanan Wi-Fi tinggi.

Karena terdapat interaksi antara Flight Distance dengan Inflight Entertainment:

Jika *Inflight_entertainment* adalah Low

- *Intercept*: 1.006.
Odds dasar kepuasan adalah $\exp(1.006) = 2.735$
- *Flight_Distance*: - 0.0004088.
Setiap kenaikan 1 km jarak penerbangan menurunkan odds kepuasan sebesar $\exp(-0.0004088) \approx 0.9996$

Jika *Inflight_entertainment* adalah Medium

- *Intercept*: 1.3464
Odds dasar kepuasan adalah $\exp(1.3464) = 3.8435$
- *Flight_Distance*: - 0.0000037
Setiap kenaikan 1 km jarak penerbangan menurunkan odds kepuasan sebesar $\exp(-0.0000037) \approx 0.9999963$

2. Berdasarkan uji klasifikasi dan kurva ROC dan dari dua model yang kita peroleh, ternyata model kedua adalah model yang terbalik. Pada model kedua ini memiliki nilai akurasi yang sudah baik dan memiliki nilai yang tinggi yaitu sebesar 86,5%,. Selain itu juga, model kedua memiliki kemampuan diskriminasi yang cukup tinggi dengan nilai AUC yaitu sebesar 0.916.

3. Faktor Peningkat dan penurun Odds Kepuasan:

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).

DAFTAR PUSTAKA

- Afifah, Devi Nur. (2020). Penerapan Metode Regresi Logistik Biner Pada Kesejahteraan Rumah Tangga Di Kabupaten Mojokerto. Malang: Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
- Basuki, Agus Tri. (2017). Regresi Logistik. Yogyakarta: Universitas Muhammadiyah
- Kaggle. (s.d.). Passenger Satisfaction Analysis. Diakses dari <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>
- Marna, Saftari, M., Jana, P., Maxrizal. (2021). Analisis Regresi Logistik Biner untuk Memprediksi Faktor Internal dan Eksternal terhadap Indeks Prestasi. Jurnal Ilmiah Pendidikan Matematika.
- S. Ghazi, R. Ramli, and A. Setyani. (2018). Analisis Keputusan Nasabah Dalam Memilih Jenis Bank: Penerapan Model Regresi Logistik Biner (Studi Kasus Pada Bank BRI Cabang Balikpapan). Media Statistika, 11(1), 17-26.
- Wahidah Alwi, E. S. (2018 vol. 6 no.1). Analisis Regresi Logistik Biner Untuk Memprediksi Kepuasan Pengunjung Pada Rumah Sakit Umum. Jurnal MSA
- Widjaja, Ezra Laurentia. (2017). Analisa Pengaruh Kualitas Layanan Terhadap Kepuasan Penumpang Maskapai Penerbangan Batik Air. Surabaya: Fakultas Ekonomi Universitas Kristen Petra.

APPENDIKS

SCRIPT R yang digunakan:

```
lokasifile <- "C:/Users/Abdul Ghofur/Downloads/Data_Planes.csv"
data <- read.csv(lokasifile)
```

```
#Eksplorasi data
str(data)
```

```
listkolom <- colnames(data)
listkolom
```

```
numerik <- c("Age",
             "Flight_Distance",
             "Departure_Delay_in_Minutes",
             "Arrival_Delay_in_Minutes"
            )
```

```
kategorik <- c("Gender",
               "Type_of_Travel",
               "Class",
               "Inflight_wifi_service",
               "Inflight_entertainment",
               "satisfaction"
              )
```

```
# Statistika Deskriptif Data Numerik
lapply(data[numerik], summary)
```

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

```
# Penyederhanaan kategori
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"
```

```
# Menyesuaikan variabel target agar menjadi biner
data$satisfaction <- ifelse(data$satisfaction == "satisfied", 1, 0)
```

```
# model dasar (menjumlahkan semua variabel yang ada
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)

# Memeriksa multikolinearitas
car::vif(model_dasar) # Cek VIF
numerik_data <- data[, sapply(data, is.numeric)]
# Korelasi untuk nilai yang lengkap
cor_matrix <- cor(numerik_data, use = "complete.obs")
cor_matrix

## Transformasi matriks korelasi ke format long
cor_melt <- melt(cor_matrix)

## Plot heatmap korelasi
ggplot(cor_melt, aes(Var1, Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "blue", high = "maroon", mid = "white",
    midpoint = 0, limit = c(-1, 1), space = "Lab",
    name = "Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  coord_fixed() +
  labs(title = "Correlation Heatmap", x = "", y = "")

# Model dasar tanpa multikolinearitas
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)

# Model Building
## Metode Stepwise Backward
model_back<-step(model_dasar2, test = "Chisq")

```

```

summary(model_back)

## Metode Stepwise 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")
summary(model_forward)

# Memeriksa apakah model sama
anova(model_back,model_forward,test = "Chisq")
## Diperoleh model sama

# Diperoleh model sederhana
model_1 <- glm(satisfaction ~
              Gender +
              Age +
              Flight_Distance +
              Departure_Delay_in_Minutes +
              Type_of_Travel +
              Class +
              Inflight_wifi_service +
              Inflight_entertainment,
              data = data, family = binomial)
summary(model_1)

# Klasifikasi tabel model pertama
yhat1 <- fitted(model_1) # prediksi peluang
pred1_50 <- ifelse(yhat1 >= 0.50, 1, 0) # prediksi 0/1
case1<-data$satisfaction

confusion_matrix1 <- table(pred1_50, case1)
confusion_matrix1 # confusion matrix

TP1 <- confusion_matrix1[2, 2] # True Positive
TN1 <- confusion_matrix1[1, 1] # True Negative

# Akurasi model pertama
akurasi1 <- (TP1 + TN1) / sum(confusion_matrix1)
print(paste("Akurasi:", akurasi1))

# Transformasi tabel menjadi data frame
confusion_long1 <- melt(confusion_matrix1)
colnames(confusion_long1) <- c("Predicted", "Observed", "Count")

# Heatmap Confusion Matrix model pertama
ggplot(confusion_long1, aes(x = Observed, y = Predicted, fill = Count)) +
  geom_tile(color = "white") +

```

```

scale_fill_gradient(low = "lightblue", high = "darkblue", name = "Count") +
geom_text(aes(label = Count), color = "white", size = 5) + # Menampilkan nilai
labs(title = "Confusion Matrix Heatmap", x = "Observed", y = "Predicted") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

# kurva ROC model pertama
roc1 <- roc(case1, yhat1)
plot.roc(roc1, main = "ROC", print.auc = TRUE, legacy.axes = TRUE)

# Model dengan 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))

# Klasifikasi tabel model interaksi
yhat2 <- fitted(model_2) # prediksi peluang
pred2_50 <- ifelse(yhat2 >= 0.50, 1, 0) # prediksi 0/1
case2 <- data$satisfaction

confusion_matrix2 <- table(pred2_50, case2)
confusion_matrix2

TP2 <- confusion_matrix2[2, 2] # True Positive
TN2 <- confusion_matrix2[1, 1] # True Negative

# Akurasi model interaksi
akurasi2 <- (TP2 + TN2) / sum(confusion_matrix2)
print(paste("Akurasi:", akurasi2))

# Transformasi tabel menjadi data frame
confusion_long2 <- melt(confusion_matrix2)
colnames(confusion_long2) <- c("Predicted", "Observed", "Count")

# Heatmap Confusion Matrix model pertama
ggplot(confusion_long2, aes(x = Observed, y = Predicted, fill = Count)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "darkblue", name = "Count") +
  geom_text(aes(label = Count), color = "white", size = 5) + # Menampilkan nilai
  labs(title = "Confusion Matrix Heatmap", x = "Observed", y = "Predicted") +
  theme_minimal() +

```



```
theme(axis.text.x = element_text(angle = 45, hjust = 1))

# kurva ROC model interaksi
roc2 <- roc(case2, yhat2)
plot.roc(roc2, main = "ROC", print.auc = TRUE, legacy.axes = TRUE)

# Membandingkan model dengan LRT
anova(model_1, model_2, test = "Chisq")

# Summary Model Akhir
summary((model_2))
```

1. Tabel Kontribusi Anggota

Nama	Deskripsi Kontribusi	Persentase
Dani Hidayat	EDA, Membuat BAB 2 dan BAB 5 Makalah, Presentasi	100%
Muhammad Fahmi Islam Fadlillah	Membuat Salindia, Membuat Abstrak dan BAB 1 Makalah, Presentasi	100%
Muhammad Abdul Ghofur	Membuat Script R, Membuat Salindia, Membuat Apendiks Makalah, Presentasi	100%
Muhammad Anthar Jailani Putra	Membuat Kerangka Makalah, Membuat BAB 2 dan BAB 4 Makalah, Presentasi	100%
Fadhilah Awal Rizqiansyah	Membuat Salindia, Membuat BAB 2 dan BAB 3 Makalah, Presentasi	100%