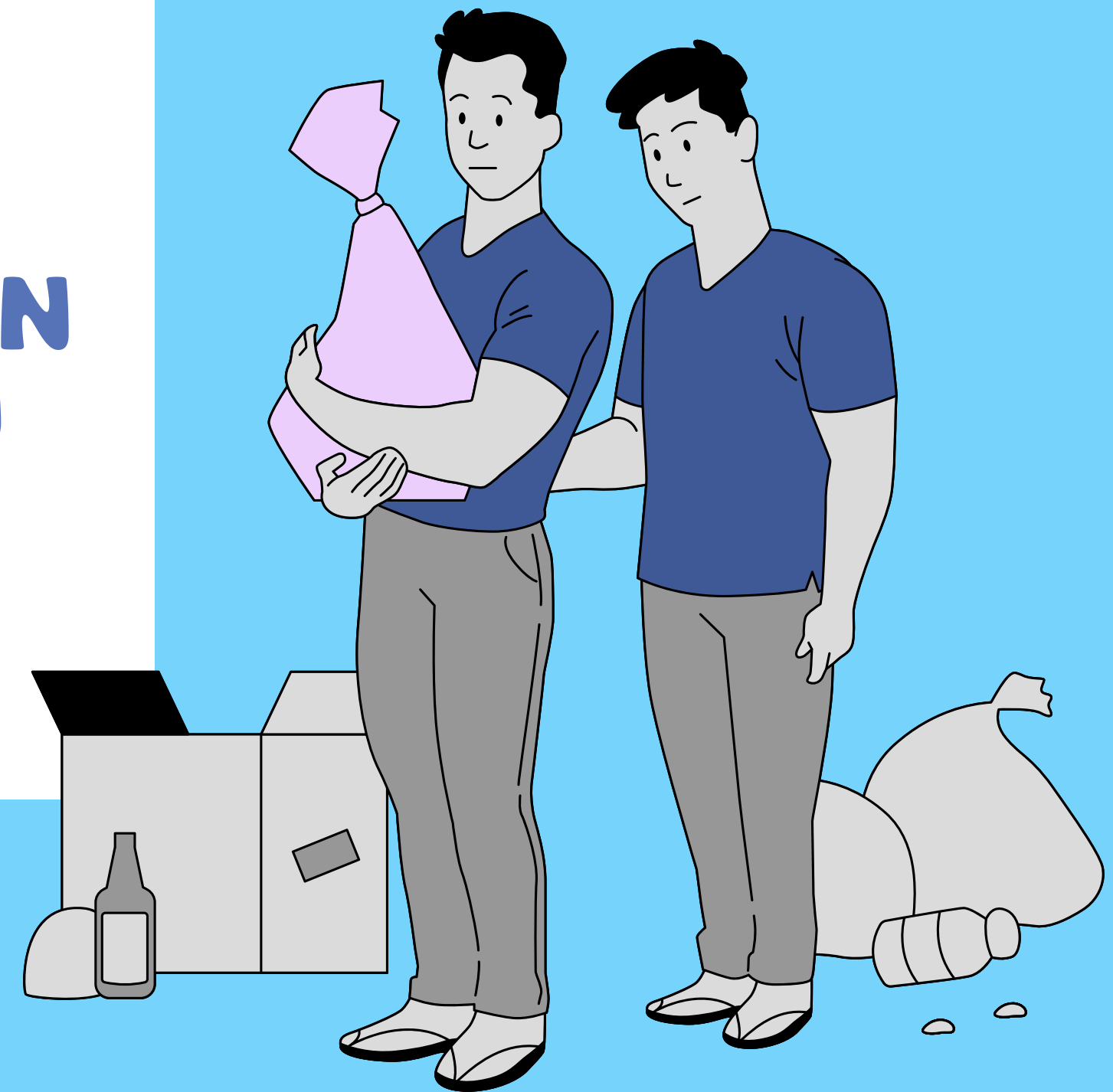




MODEL REGRESI PROBIT SEBAGAI PEMODELAN PERSENTASE PENDUDUK MISKIN KABUPATEN / KOTA DI PULAU JAWA

Kelompok 8 Kelas Ekonometri (B)



ANGGOTA KELOMPOK



Dwight J.O. Hutagalung
2006571160



Owen Susanto
2006570933



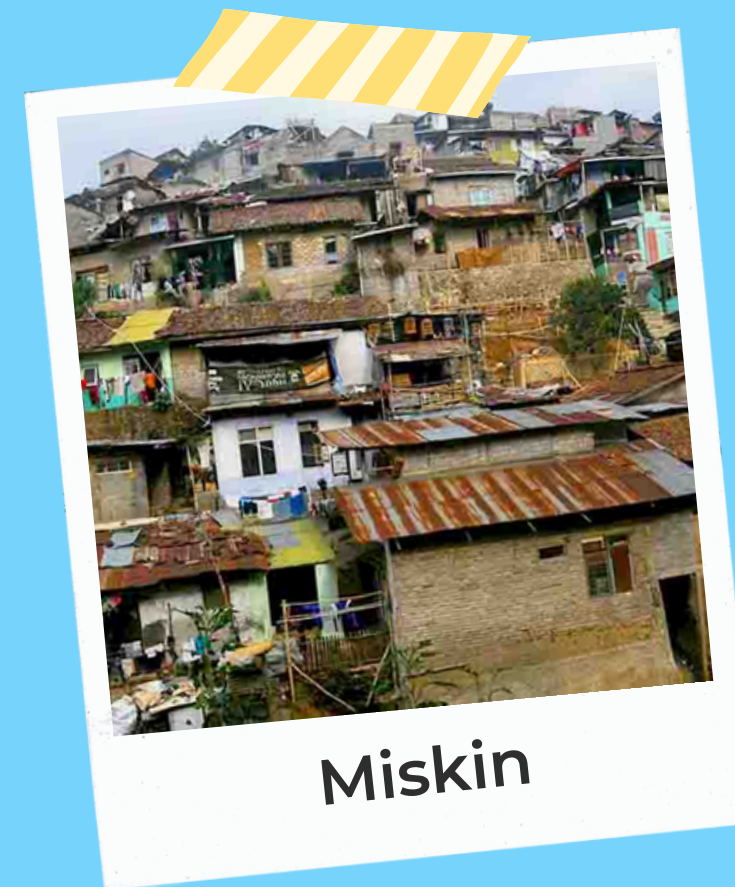
Evan Haryowidyatna
2006485011

6

LATAR BELAKANG

Daerah Miskin harus menjadi target utama pembangunan

Hal ini sesuai dengan target pemerintah Indonesia untuk pemeratakan seluruh daerah agar memiliki kekuatan ekonomi yang sama

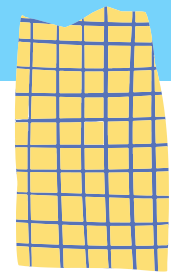




“

**If you define the
problem correctly, you
almost have the
solution**

**Steve
Jobs**



RUMUSAN MASALAH

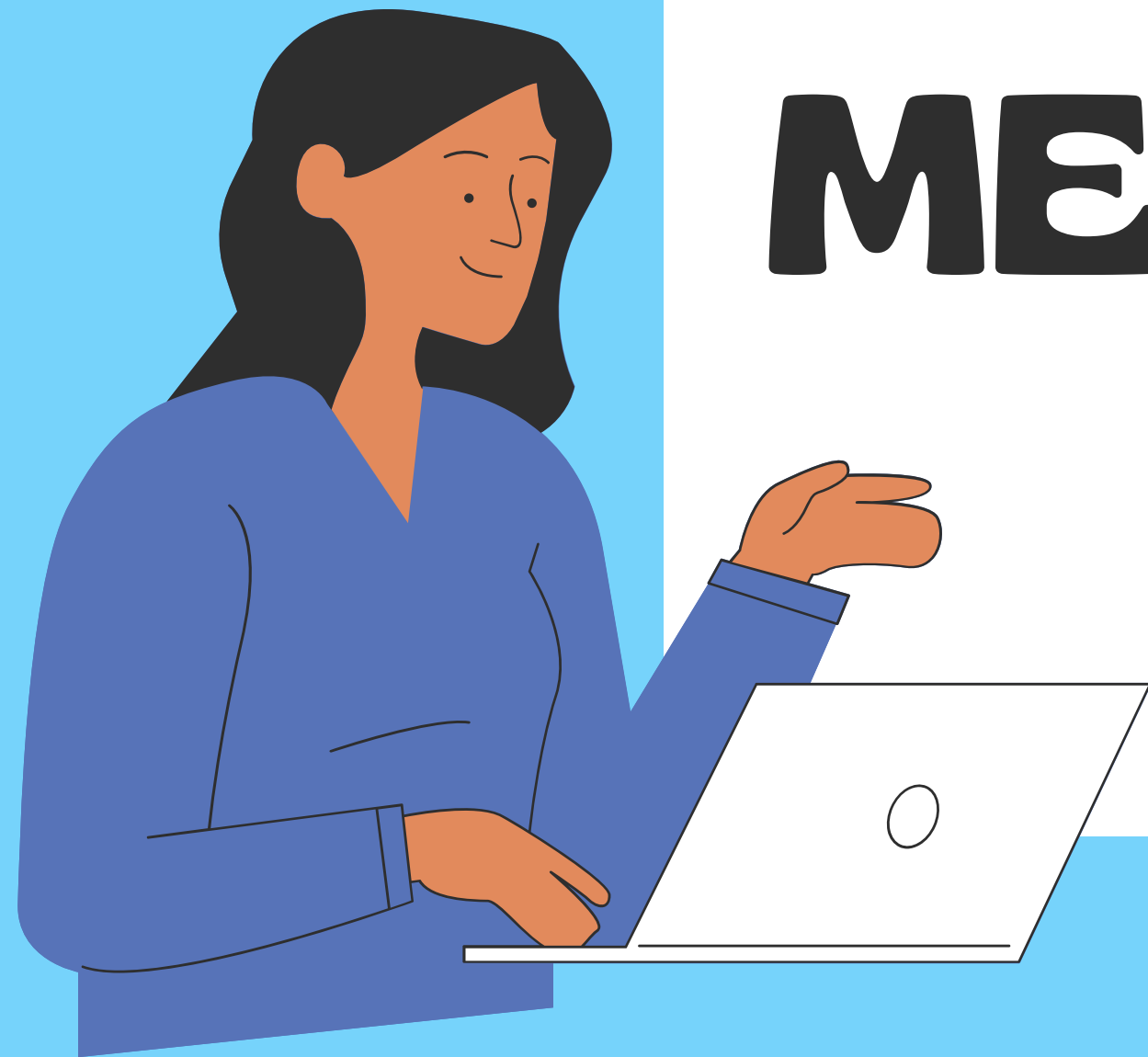
- Bagaimana cara memperoleh taksiran parameter dengan model regresi probit ?
- Variabel apa yang memiliki pengaruh signifikan pada probabilitas daerah dengan ekonomi di bawah rata-rata?

Tujuan Penelitian

Menguji variabel yang signifikan untuk probabilitas daerah dengan ekonomi di bawah rata-rata

Membuat model yang berpengaruh secara signifikan terhadap masalah.

Mengidentifikasi cara memperoleh taksiran parameter dari model probit.



METODOLOGI

METODOLOGI

UJI MULTIKOLINEARITAS

Menguji multikolinearitas dari variabel bebas dengan melihat nilai dari VIF.

- Jika **VIF** < 10, maka tidak terdapat multikolinearitas.
- Jika **VIF** > 10, maka terdapat multikolinearitas.

MODEL REGRESI PROBIT

Membuat model regresi probit dengan meregresikan variabel target (Y), yaitu apakah persentase penduduk miskin di atas rata-rata, terhadap variabel - variabel prediktornya (X1 hingga X6) di mana parameter model diestimasi dengan menggunakan metode Maximum Likelihood Estimation.

EVALUASI KETEPATAN MODEL (APER)

Melakukan evaluasi ketepatan klasifikasi model yang dilakukan dengan melihat peluang kesalahan klasifikasi model atau APER (Apparent Error Rate) dimana formulanya adalah sebagai berikut.

$$1 - APER = 1 - \left(\frac{n_{12} + n_{21}}{N} \right) \times 100\%$$

METODOLOGI

- 1 Memodelkan faktor - faktor yang menjadi variabel prediktor dengan regresi probit dan memeriksa variabel mana saja yang signifikan
- 2 Membuat model baru dengan menghapus variabel yang tidak signifikan guna memilih model terbaik
- 3 Melakukan uji asumsi untuk mengetahui apakah model sudah memenuhi semua syarat yang harus dipenuhi
- 4 Membuat confusion matrix untuk memeriksa akurasi model dan dilanjutkan dengan menginterpretasi model



HASIL ANALISIS



DATA

Data didapatkan dari BPS.go.id

Data dikumpulkan dengan melakukan scraping dari beberapa sumber tabel, dengan variabel sebagai berikut :



VARIABEL DATA

| Simbol | Nama Variabel | Tipe | Keterangan |
|--------|---------------------------------------|-----------|---|
| Y | Daerah Miskin | Kategorik | 1 - Daerah dengan persentase penduduk miskin di atas rata-rata Pulau Jawa. 0 - Daerah dengan persentase penduduk miskin di bawah rata-rata Pulau Jawa. |
| X1 | Provinsi | Kategorik | Pulau Jawa |
| X2 | Garis Kemiskinan | Numerik | Pulau Jawa dalam (Rp) |
| X3 | Kepadatan Penduduk | Numerik | Pulau Jawa dalam (Jiwa) |
| X4 | Persentase Kepemilikan Rumah Sendiri | Numerik | Pulau Jawa |
| X5 | Persentase Jiwa Dengan Akses Internet | Numerik | Pulau Jawa |
| X6 | Pengeluaran per Kapita | Numerik | Pulau Jawa dalam (Rp) |

SUMMARY MODEL 1

1 Pengujian Serentak

Untuk melakukan pengujian serentak, akan dilihat dari *null* dan *residual deviance*, di mana $X^2 = \text{null deviance} - \text{residual deviance} = 164.961 - 77.301 = 87.66 > \text{Chi Square}(0.05, 10) = 18.307$, sehingga paling tidak terdapat satu variabel yang signifikan di dalam model.

Selain itu, apabila kita menggunakan *chi-square to p-value calculator*, didapatkan bahwa *p-value* sebesar $0,000 < 0,05$, sehingga dapat disimpulkan bahwa model sudah cukup baik atau berguna untuk memprediksi probabilitas sebuah daerah memiliki persentase masyarakat miskin di atas atau bawah rata-rata.

2 Pengujian Parsial

Untuk melakukan pengujian parsial, dapat dilihat dari *p-value* dengan melihat lambang (*) di bagian kanan variabel, di mana lambang tersebut melambangkan variabel yang signifikan terhadap model atau $p\text{-value} < 0.05$ untuk variabel tersebut. Di model ini, terdapat beberapa variabel yang tidak signifikan.

```
> summary(model)
```

```
Call:
glm(formula = Binary ~ provinsi + garis_kemiskinan + pengeluaran_perkapita +
    milik_rumah + kepadatan_penduduk + akses_internet, family = binomial(link = "probit"),
    data = data)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2292 -0.5594  0.0000  0.2480  2.1968
```

```
Coefficients:
```

| | Estimate | Std. Error | z value | Pr(> z) |
|-----------------------|------------|------------|---------|------------|
| (Intercept) | -6.195e+00 | 4.697e+00 | -1.319 | 0.18720 |
| provinsiDI Yogyakarta | -4.080e+00 | 1.498e+00 | -2.724 | 0.00645 ** |
| provinsiDKI Jakarta | -4.561e+00 | 4.234e+01 | -0.108 | 0.91422 |
| provinsiJawa Barat | -7.686e-01 | 9.718e-01 | -0.791 | 0.42897 |
| provinsiJawa Tengah | -2.667e+00 | 1.025e+00 | -2.603 | 0.00925 ** |
| provinsiJawa Timur | -2.553e+00 | 1.028e+00 | -2.483 | 0.01302 * |
| garis_kemiskinan | -1.048e-05 | 4.815e-06 | -2.176 | 0.02957 * |
| pengeluaran_perkapita | 2.305e-07 | 1.862e-07 | 1.238 | 0.21557 |
| milik_rumah | 5.310e-02 | 4.323e-02 | 1.228 | 0.21936 |
| kepadatan_penduduk | 6.429e-04 | 2.692e-04 | 2.388 | 0.01694 * |
| akses_internet | 8.072e-02 | 3.646e-02 | 2.214 | 0.02685 * |

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```


```
Null deviance: 164.961  on 118  degrees of freedom
Residual deviance:  77.301  on 108  degrees of freedom
AIC: 99.301
```

```
Number of Fisher Scoring iterations: 13
```

untuk uji parsial

untuk uji serentak

PEMILIHAN MODEL TERBAIK



Dari *summary* model 1 yang telah diberikan sebelumnya, memang didapatkan bahwa variabel - variabel yang digunakan dapat membentuk model yang baik. Akan tetapi, kelompok kami akan mengeluarkan variabel - variabel yang tidak signifikan (dapat terlihat dari *p-value* masing-masing variabel). Lalu, kami akan melihat model mana yang paling baik dengan melihat kriteria AIC yang dikeluarkan.

SUMMARY MODEL 2

1 Pengujian Serentak

Untuk melakukan pengujian serentak, akan dilihat dari *null* dan *residual deviance*, di mana $X^2 = \text{null deviance} - \text{residual deviance} = 164.961 - 80.052 = 84.909 > \text{Chi Square}(0.05, 8) = 15.51$, sehingga paling tidak terdapat satu variabel yang signifikan di dalam model.

Selain itu, apabila kita menggunakan *chi-square to p-value calculator*, didapatkan bahwa *p-value* sebesar $0,000 < 0,05$, sehingga dapat disimpulkan bahwa model sudah cukup baik atau berguna untuk memprediksi probabilitas sebuah daerah memiliki persentase masyarakat miskin di atas atau bawah rata-rata.

2 Pengujian Parsial

Untuk melakukan pengujian parsial, dapat dilihat dari *p-value* dengan melihat lambang (*) di bagian kanan variabel, di mana lambang tersebut melambangkan variabel yang signifikan terhadap model atau *p-value* < 0.05 untuk variabel tersebut. Dalam model kedua ini, seluruh variabel signifikan terhadap model.

```
Call:
glm(formula = Binary ~ provinsi + garis_kemiskinan + kepadatan_penduduk +
    akses_internet, family = binomial(link = "probit"), data = data)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -2.4720 | -0.5607 | 0.0000 | 0.2834 | 2.2181 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-----------------------|------------|------------|---------|------------|
| (Intercept) | -9.069e-02 | 1.506e+00 | -0.060 | 0.95199 |
| provinsiDI Yogyakarta | -3.741e+00 | 1.356e+00 | -2.759 | 0.00580 ** |
| provinsiDKI Jakarta | -3.538e+00 | 1.230e+01 | -0.288 | 0.77358 |
| provinsiJawa Barat | -1.071e+00 | 9.034e-01 | -1.185 | 0.23582 |
| provinsiJawa Tengah | -2.379e+00 | 9.170e-01 | -2.594 | 0.00949 ** |
| provinsiJawa Timur | -2.159e+00 | 9.163e-01 | -2.356 | 0.01846 * |
| garis_kemiskinan | -8.912e-06 | 4.462e-06 | -1.997 | 0.04579 * |
| kepadatan_penduduk | 4.478e-04 | 2.015e-04 | 2.222 | 0.02627 * |
| akses_internet | 9.342e-02 | 2.957e-02 | 3.159 | 0.00158 ** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 164.961 on 118 degrees of freedom
Residual deviance: 80.052 on 110 degrees of freedom
AIC: 98.052

Selain itu, karena AIC dari model 2 (sebesar 98.052) lebih kecil daripada AIC dari model 1 (sebesar 99.031), maka kami akan memilih untuk menggunakan model 2.

untuk uji parsial

untuk uji serentak

UJI MULTIKOLINEARITAS

Didapatkan, seluruh variabel memiliki nilai $VIF < 10$.

Karena $VIF < 10$, maka mengindikasikan bahwa tidak terjadi multikolinearitas pada semua variabel.

Selain itu, diambil berdasarkan salah satu publikasi dari Macquaire University, Australia, didapatkan *rule of thumb* kolinearitas dari $GVIF^{(1/(2*Df))}$ adalah lebih besar 2.5, di mana dari model ini terlihat bahwa tidak ada variabel dengan multikolinearitas.

```
> vif(modelbaru)
```

| | GVIF | Df | $GVIF^{(1/(2*Df))}$ |
|--------------------|----------|----|---------------------|
| provinsi | 1.533615 | 5 | 1.043690 |
| garis_kemiskinan | 2.165213 | 1 | 1.471466 |
| kepadatan_penduduk | 1.966294 | 1 | 1.402246 |
| akses_internet | 2.560178 | 1 | 1.600056 |

MODEL REGRESI PROBIT

Dari uji parameter yang telah dilakukan pada slide sebelumnya, berikut ini didapatkan model regresi probit biner dengan persamaan probabilitas suatu kabupaten/kota di Pulau Jawa memiliki persentase penduduk miskin di atas rata-rata.

$$\begin{aligned}\Pr(Y = 1) &= \Phi(Z) \\ &= \Phi(-0.09069 - 3.741DIY - 3.538DKI - 1.071Jabar - 2.379Jateng \\ &\quad - 2.159Jatim - 0.000008912GK + 0.0004478KP + 0.09342AI)\end{aligned}$$

Dengan DIY, DKI, Jabar, Jateng, dan Jatim merupakan nama-nama provinsi, GK adalah garis kemiskinan KP adalah kepadatan penduduk, dan AI adalah akses daerah terhadap internet.

EVALUASI MODEL

1 Nilai APER (Apparent Error Rate)

$$APER = \frac{12 + 7}{52 + 12 + 7 + 48} = \frac{19}{119} = 0.159$$

$(1 - APER) * 100\% = 84 \%$

2 Ketepatan Akurasi

Dari nilai APER yang ada, didapatkan ketepatan ketepatan akurasi sebesar 84%

Confusion
Matrix

| Actual // Predicted | Y=1 | Y=0 |
|---------------------|-----|-----|
| Y=1 | 52 | 12 |
| Y=0 | 7 | 48 |

KESIMPULAN & SARAN

1

Terdapat beberapa variabel yang tidak signifikan sehingga kami menghapusnya dari model berdasarkan pemilihan model terbaik dengan membandingkan AIC kedua model

2

Nilai ketepatan klasifikasi didapatkan sebesar 84% dengan nilai kesalahan klasifikasi 16%.

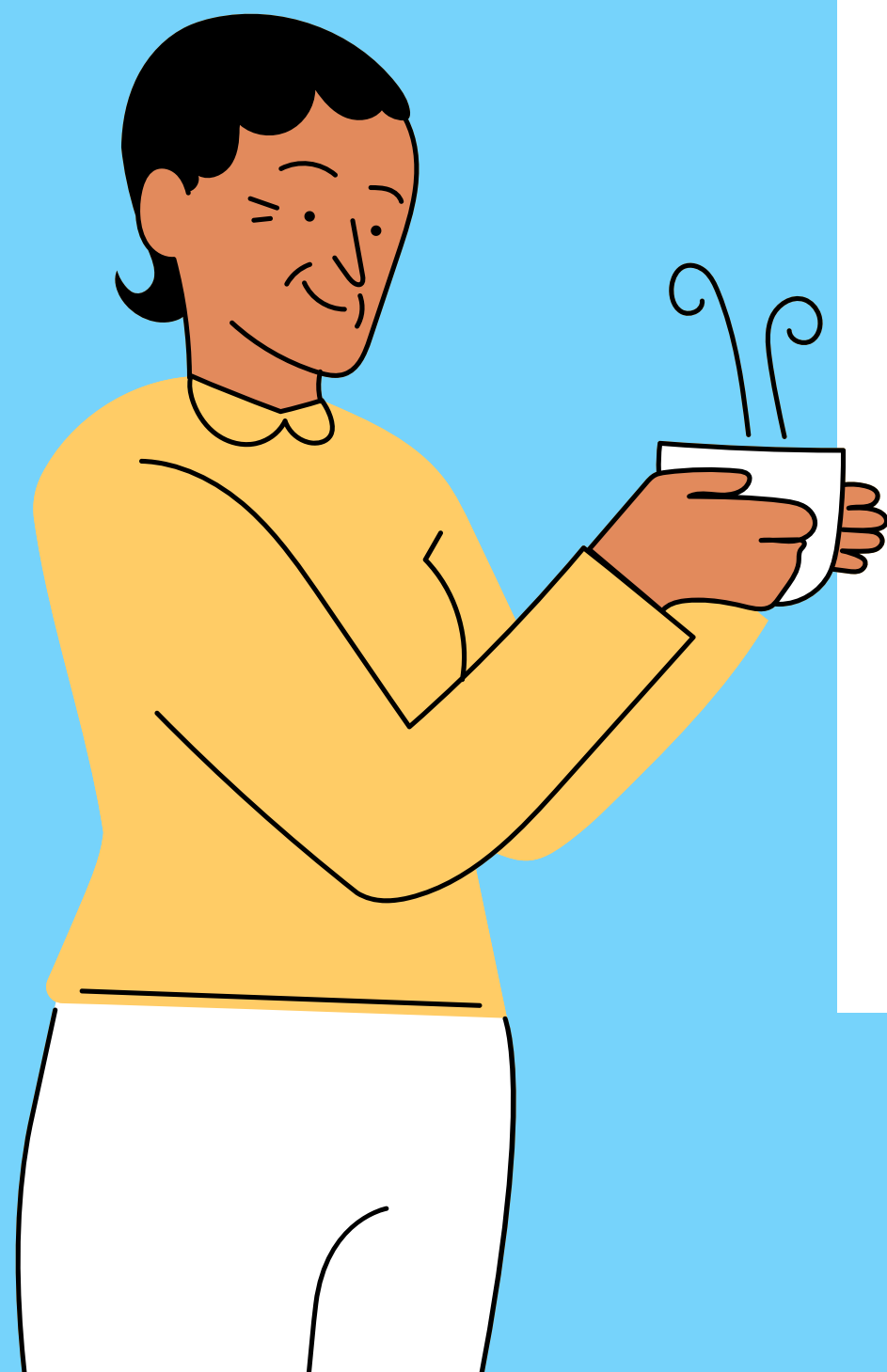


3

Menggunakan regresi probit, dapat diprediksi dengan cukup baik daerah dengan persentase penduduk miskin di atas rata-rata, tetapi perlu diingat bahwa variabel yang dihapus juga memiliki peran penting dan sebetulnya dapat mempengaruhi hasil penilaian persentase penduduk miskin suatu kabupaten kota

4

Sebaiknya, data yang diambil lebih banyak lagi dengan jumlah variabel prediktor yang lebih baik karena pastinya masih ada banyak faktor lain yang dapat mempengaruhi persentase penduduk miskin suatu daerah



**THANK
YOU!**

Have a great
day ahead.