

**Laporan Resmi Regresi Logistik Biner
Berdasarkan Indeks Pembangunan Manusia (IPM) Provinsi
Maluku Menggunakan Data Tahun 2019 - 2022**



Disusun Oleh:

Aurelia Hapsari Dyah Rinjani	3323600035
Evinda Eka Ayudia Lestari	3323600039
Robi'Atul Adawiyah	3323600041
Wahyu Ikbal Maulana	3323600056

**Mata Kuliah Praktikum Pemodelan Statistik Terapan
Program Studi D4 Sains Data Terapan
Departemen Teknik Informatika dan Komputer
Politeknik Elektronika Negeri Surabaya
2024**

LEMBAR PENGESAHAN

Laporan resmi ini untuk memenuhi tugas mata kuliah praktikum pemodelan statistik terapan yang disusun oleh.

1. Nama : Aurelia Hapsari Dyah Rinjani
NRP : 3323600035
2. Nama : Evinda Eka Ayudia Lestari
NRP : 3323600039
3. Nama : Robi' Atul Adawiyah
NRP : 3323600041
4. Nama : Wahyu Iqbal Maulana
NRP : 3323600056

Surabaya, Maret 2024
Mengetahui,
Dosen Pengampu

Ronny Susetyoko, S.Si., M.Si
NIP. 197112111995011001

ABSTRAK

Indonesia terdiri dari 38 provinsi yang tersebar mulai dari Sabang hingga Merauke, salah satunya adalah provinsi Maluku. Di Provinsi Maluku, seperti halnya di banyak daerah di Indonesia, peningkatan IPM dianggap sebagai bagian penting dalam upaya pembangunan yang berkelanjutan untuk mewujudkan kesejahteraan masyarakat. Analisis terkait faktor apa saja yang mempengaruhi IPM dapat menggunakan metode regresi logistik. Regresi logistik adalah salah satu bentuk regresi dimana variabel responnya berbentuk biner. Data yang digunakan dalam analisis ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Maluku menggunakan data tahun 2019-2022. Metodologi pemodelan regresi logistik dalam praktikum ini dimulai dari penetapan kerangka masalah, pengumpulan data, data preprocessing, membagi data training dan data testing, pemodelan dan pengujian, evaluasi kinerja model, dan seleksi model klasifikasi. Secara keseluruhan, hasil dari model regresi logistik biner telah menunjukkan kinerja yang sangat baik dan dapat digunakan sebagai alat yang efektif untuk memprediksi IPM sedang atau tinggi. ROC AUC model sebesar 0.89 berarti, model klasifikasi logit biner bekerja dengan sangat baik memprediksi IPM sedang atau tinggi.

Kata kunci: IPM, Logistik biner, Regresi

DAFTAR ISI

LEMBAR PENGESAHAN.....	2
ABSTRAK.....	3
DAFTAR ISI.....	4
DAFTAR TABEL DAN GAMBAR.....	5
BAB 1 PENDAHULUAN.....	6
1.1. Latar Belakang.....	6
1.2. Tujuan.....	7
1.3. Rumusan Masalah.....	7
BAB 2 TINJAUAN PUSTAKA.....	8
2.1. Indeks Pembangunan Manusia.....	8
2.2. Persentase Penduduk Miskin.....	8
2.3. Angka Partisipasi Kasar SMA/ sederajat.....	9
2.4. Persentase Tingkat Pengangguran Terbuka.....	9
2.5. Persentase Penduduk yang Mengalami Keluhan Kesehatan.....	9
BAB 3 METODE PENELITIAN.....	11
3.1. Dataset.....	11
3.2. Metodologi.....	11
3.3. Algoritma.....	12
3.3.1. Regresi Linier.....	12
3.3.2. Regresi Logistik Biner.....	12
BAB 4 HASIL PENELITIAN DAN LUARAN YANG DICAPAI.....	14
4.1. Dataset yang digunakan.....	14
4.2. Exploratory Data Analysis (EDA).....	14
4.3. Logistic Model.....	15
4.4. Evaluasi model.....	18
BAB 5 KESIMPULAN DAN SARAN.....	21
5.1. Interpretasi Hasil.....	21
5.2. Kesimpulan.....	21
5.3. Saran.....	22
DAFTAR PUSTAKA.....	23

DAFTAR TABEL DAN GAMBAR

Tabel 1. Dataset.....	14
Gambar 1. Boxplot.....	14
Gambar 2. Histogram.....	15
Gambar 3. Plot ROC Kurva.....	19
Gambar 4. Confusion Matrix.....	19

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Indonesia terdiri dari 38 provinsi yang tersebar mulai dari Sabang hingga Merauke, dengan masing-masing provinsi memiliki ibukota di wilayahnya sendiri. Salah satunya adalah provinsi Maluku, yang memiliki Kota Ambon sebagai ibu kotanya. Provinsi Maluku terbagi menjadi 9 kabupaten dan 2 kotamadya. Di setiap provinsi perlu adanya pembangunan berkelanjutan dalam upaya mewujudkan kesejahteraan masyarakat di provinsi tersebut. Salah satu indikator yang biasanya dijadikan sebagai perhatian utama dalam mewujudkan kesejahteraan masyarakat adalah Indeks Pembangunan Manusia (IPM).

Indeks Pembangunan Manusia (IPM) adalah sebuah alat ukur yang digunakan untuk mengevaluasi kualitas pembangunan manusia dari berbagai aspek, baik yang bersifat fisik maupun non-fisik. Dalam konteks fisik, hal ini mencakup kesehatan dan kesejahteraan, yang dapat tercermin dalam angka harapan hidup dan daya beli masyarakat. Sementara itu, aspek non-fisik melibatkan kualitas pendidikan masyarakat. IPM memberikan gambaran yang luas mengenai kinerja pembangunan suatu wilayah dengan memperhatikan aspek-aspek penting seperti harapan hidup, pendidikan, dan standar hidup layak. IPM mencerminkan kemajuan suatu wilayah dalam memenuhi kebutuhan dasar penduduknya. Di Provinsi Maluku, seperti halnya di banyak daerah di Indonesia, peningkatan IPM dianggap sebagai bagian penting dari upaya pembangunan yang berkelanjutan.

Oleh karena itu, penting dilakukan analisis terkait faktor apa saja yang mempengaruhi IPM. Salah satu cara yang dapat digunakan adalah metode regresi logistik. Analisis regresi adalah analisis sederhana untuk mengetahui hubungan fungsional antara variabel. Hubungan tersebut dinyatakan dalam suatu model atau persamaan yang menghubungkan antara variabel respon dan variabel prediktor. Salah satu jenis regresi yang sering dilakukan adalah regresi logistik. Regresi logistik adalah salah satu bentuk regresi dimana variabel responnya berbentuk biner. Hal ini berarti nilai dari variabel responnya hanya ada 2 kategori yaitu $Y = 1$ (sukses) dan $Y = 0$ (gagal), sehingga pada penelitian ini diaplikasikan konsep regresi logistik biner.

1.2. Tujuan

1. Menganalisis pengaruh faktor-faktor sosial dan ekonomi Provinsi Maluku yang berpengaruh terhadap IPM pada tahun 2019-2022.
2. Mengidentifikasi variabel prediktor yang signifikan.
3. Mengetahui hubungan antara variabel prediktor (persentase penduduk miskin, APK SMA/SMK, persentase penduduk miskin, persentase tingkat pengangguran terbuka, persentase penduduk yang mengalami keluhan kesehatan) dengan variabel Indeks Pembangunan Manusia pada tahun 2019-2022.

1.3. Rumusan Masalah

1. Faktor apa saja yang memiliki pengaruh signifikan terhadap kategori IPM provinsi Maluku di rentang waktu 2019-2022?
2. Bagaimana kinerja model prediksi logistik biner dalam memprediksi kategori IPM provinsi Maluku di rentang waktu 2019-2022?
3. Bagaimana kontribusi masing-masing variabel prediktor terhadap kategori IPM provinsi Maluku di rentang waktu 2019-2022?
4. Bagaimana variabilitas data mempengaruhi ketepatan prediksi kategori IPM provinsi Maluku di rentang waktu 2019-2022?

BAB 2

TINJAUAN PUSTAKA

2.1. Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) suatu indikator yang dapat menggambarkan perkembangan pembangunan manusia secara terukur dan representatif (UNDP, 1990). IPM merupakan indeks komposit yang dihitung sebagai rata-rata sederhana dari tiga indeks dari dimensi yang menggambarkan kemampuan dasar manusia dalam memperluas pilihan-pilihan. Rumus umum yang dipakai adalah sebagai berikut (UNDP, 2004).

$$IPM = \frac{1}{3} (Y_1 + Y_2 + Y_3)$$

Dengan penjelasan:

IPM = Indeks Pembangunan Manusia

Y1 = Indeks Harapan Hidup

Y2 = Indeks Pendidikan

Y3 = Indeks Standard Hidup Layak

2.2. Persentase Penduduk Miskin

Menurut (Yacoub, 2012) dalam penelitiannya menyatakan bahwa kemiskinan merupakan salah satu persoalan mendasar, karena kemiskinan menyangkut pemenuhan kebutuhan yang paling mendasar dalam kehidupan dan kemiskinan merupakan masalah global karena kemiskinan merupakan masalah yang dihadapi banyak negara. Menurut (World Bank, 2004) salah satu sebab kemiskinan adalah karena kurangnya pendapatan dan aset (lack of income and assets) untuk memenuhi kebutuhan dasar seperti makanan, pakaian, perumahan, tingkat kesehatan dan pendidikan yang dapat diterima (acceptable). Di samping itu kemiskinan juga berkaitan dengan keterbatasan lapangan pekerjaan dan biasanya mereka yang dikategorikan miskin (the poor) tidak memiliki pekerjaan (pengangguran), serta tingkat pendidikan dan kesehatan mereka pada umumnya tidak memadai.

Persentase Penduduk Miskin (Headcount Index/P0) adalah Persentase penduduk miskin yang berada di bawah garis kemiskinan. Headcount Index secara sederhana mengukur proporsi yang dikategorikan miskin. Mengetahui persentase penduduk yang

dikategorikan miskin. Angka yang ditunjukkan oleh HCI-P0 menunjukkan proporsi penduduk miskin di suatu wilayah. Persentase penduduk miskin yang tinggi menunjukkan bahwa tingkat kemiskinan di suatu wilayah juga tinggi.

2.3. Angka Partisipasi Kasar SMA/ sederajat

Angka partisipasi kasar (APK) adalah persentase perbandingan antara jumlah penduduk yang masih bersekolah di jenjang pendidikan tertentu (tanpa memandang usia penduduk tersebut) dengan jumlah penduduk yang memenuhi syarat resmi penduduk usia sekolah di jenjang pendidikan yang sama. Dari pengertian tersebut dapat disimpulkan bahwa APK SMA/ sederajat di Maluku adalah perbandingan antara jumlah penduduk yang masih bersekolah di jenjang pendidikan SMA/ sederajat di Maluku dengan jumlah penduduk usia 16-18 tahun di Maluku dikali seratus persen. Secara perhitungan dapat ditulis $J = \frac{K}{L} \times 100\%$ dengan J adalah APK SMA/ sederajat, K adalah jumlah murid SMA/ sederajat, dan L adalah Jumlah penduduk usia 16-18 tahun di Indonesia.

Secara umum APK merupakan indikator pendidikan, artinya APK digunakan sebagai tolak ukur kesuksesan program pendidikan yang dicanangkan pemerintah dalam upaya memberikan kesempatan yang lebih luas bagi penduduk untuk mengenyam pendidikan. Merangkum dari hasil penelitian sebelumnya, secara umum faktor-faktor yang memengaruhi angka partisipasi kasar (APK) diantaranya jumlah sekolah, persentase penduduk miskin, anggaran pendidikan, serta rasio murid dan guru.

2.4. Persentase Tingkat Pengangguran Terbuka

Pengangguran adalah orang yang masuk dalam usia angkatan kerja namun belum mendapat kesempatan bekerja tetapi sedang mencari pekerjaan atau tidak. Tingkat Pengangguran Terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Dimana, TPT memiliki kegunaan dalam mengindikasikan besarnya persentase angkatan kerja yang termasuk dalam pengangguran (BPS, 2015). Adapun dalam perhitungannya digunakan rumus sebagai berikut.

$$TPT = \frac{\text{Jumlah Penganggur}}{\text{Jumlah Angkatan Kerja}} \times 100\%$$

2.5. Persentase Penduduk yang Mengalami Keluhan Kesehatan

Keluhan kesehatan adalah gangguan terhadap kondisi fisik maupun jiwa, termasuk karena kecelakaan, atau hal lain yang menyebabkan terganggunya kegiatan sehari-hari. Pada umumnya keluhan kesehatan utama yang banyak dialami oleh penduduk adalah

panas, sakit kepala, batuk, pilek, diare, asma/sesak nafas, sakit gigi. Orang yang menderita penyakit kronis dianggap mempunyai keluhan kesehatan walaupun pada waktu survei (satu bulan terakhir) yang bersangkutan tidak kambuh penyakitnya. Ada beberapa upaya yang dapat kita lakukan dalam menanggulangi masalah keluhan kesehatan seperti peningkatan gizi, penambahan fasilitas kesehatan, pelaksanaan imunisasi, penyediaan pelayanan kesehatan gratis, pengadaan obat generik, penambahan jumlah tenaga medis, dan melakukan penyuluhan tentang arti pentingnya kebersihan dan pola hidup sehat.

Sehingga dalam hal ini, persentase penduduk yang mengalami keluhan kesehatan adalah persentase jumlah penduduk yang mengalami keluhan kesehatan gangguan terhadap kondisi fisik maupun jiwa, termasuk karena kecelakaan, atau hal lain yang menyebabkan terganggunya kegiatan sehari-hari.

BAB 3

METODE PENELITIAN

3.1. Dataset

Data yang digunakan dalam analisis ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) menggunakan data tahun 2019-2022. Dataset ini mencakup berbagai variabel berdasarkan Indeks Pembangunan Manusia (IPM) Provinsi Maluku dengan variabel yang digunakan dalam analisis ini adalah sebagai berikut:

x1 = Persentase penduduk miskin

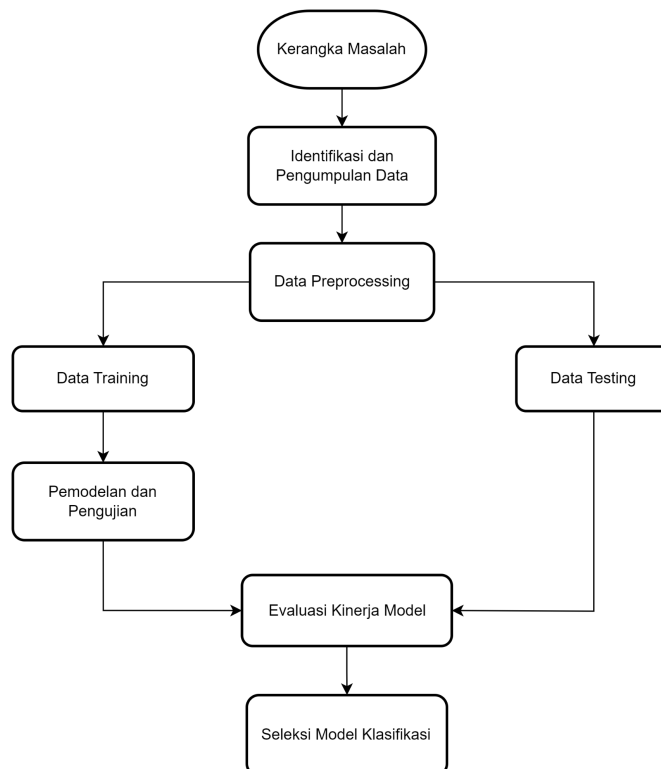
x2 = Angka partisipasi kasar SMA/SMK

x3 = Persentase tingkat pengangguran terbuka

x4 = Persentase penduduk yang mengalami keluhan kesehatan

3.2. Metodologi

Metodologi pemodelan regresi logistik dalam praktikum ini dimulai dari penetapan kerangka masalah, pengumpulan data, data preprocessing, membagi data training dan data testing, pemodelan dan pengujian, evaluasi kinerja model, dan seleksi model klasifikasi.



3.3. Algoritma

3.3.1. Regresi Linier

Regresi linear sederhana adalah persamaan regresi yang menggambarkan hubungan antara satu variabel bebas (X) dan satu variabel tak bebas (Y), dimana hubungan keduanya dapat digambarkan sebagai suatu garis lurus. Persamaan regresi dapat terdiri dari satu atau lebih variabel bebas dan satu variabel terikat. Persamaan yang terdiri dari satu variabel bebas dan satu variabel terikat disebut persamaan regresi sederhana, sedangkan yang terdiri dari satu variabel terikat dan beberapa variabel bebas disebut persamaan regresi berganda. Hubungan kedua variabel X dan Y dapat dituliskan dalam bentuk persamaan:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \dots\dots\dots$$

dengan keterangan sebagai berikut:

Y = Variabel tak bebas

X = Variabel bebas

β_0 = intersep/perpotongan dengan sumbu tegak

β_1 = Kemiringan/gradien, ε_1 error yang saling bebas dan menyebar normal $N(0, \sigma^2)$ $i = 1, 2, \dots, n$.

3.3.2. Regresi Logistik Biner

Menurut Harlan dalam (Brahmantyo et al., 2021), Regresi Logistik Biner digunakan untuk mengukur pengaruh variabel independen kontinu atau kategoris (X) terhadap dependen variabel (Y) yang memiliki nilai dikotomis atau biner. Bentuk umum regresi logistik:

$$\begin{aligned} \text{logit}(Y) &= \ln O(Y) = \ln \frac{P(Y)}{1 - P(Y)} \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \end{aligned} \quad (1)$$

yang mana dapat ditulis dalam bentuk Odds dari Y dengan:

$$O(Y) = \frac{P(Y)}{1 - P(Y)} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p} \quad (2)$$

maka dari persamaan (2) dapat dibentuk menjadi persamaan probabilistik dalam regresi logistik:

$$P(Y) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}} \quad (3)$$

di mana persamaan (3) dapat digunakan untuk menghitung nilai probabilitas dalam model regresi logistik. Estimasi parameter regresi logistik untuk variabel tak bebas yang dikotomis atau biner, menggunakan Distribusi Bernoulli sebagai distribusi probabilitas. Ini akan menghasilkan fungsi likelihood yang merupakan kombinasi dari semua parameter fungsi probabilitas dalam model. Oleh karena itu, dapat dikatakan bahwa fungsi likelihood untuk regresi logistik adalah:

$$L(\beta) = \prod_{i=1}^n [P(Y)]^{Y_i} \cdot [1 - P(Y)]^{1 - Y_i}, Y_i = \{0, 1\} \quad (4)$$

maka ketika persamaan (3) disubstitusikan ke persamaan (4), maka akan menghasilkan:

$$L(\beta) = \prod_{i=1}^n [e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}}]^{Y_i} \cdot [1 + e^{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}}] \quad (5)$$

dan jika persamaan (5) ditransformasikan dengan logaritma natural, menjadi

$$\ln L(\beta) = \sum_{i=1}^n \left[Y_i \left\{ \sum_{p=0}^n \beta_p X_{ip} - \ln \left(1 + e^{\sum_{p=0}^n \beta_p X_{ip}} \right) \right\} \right] \quad (6)$$

Adapun pada persamaan (6) akan diestimasi dengan pendekatan Newton Raphson, karena fungsi tidak dalam bentuk linier dan perlu pendekatan untuk memperkirakan nilai parameter

BAB 4

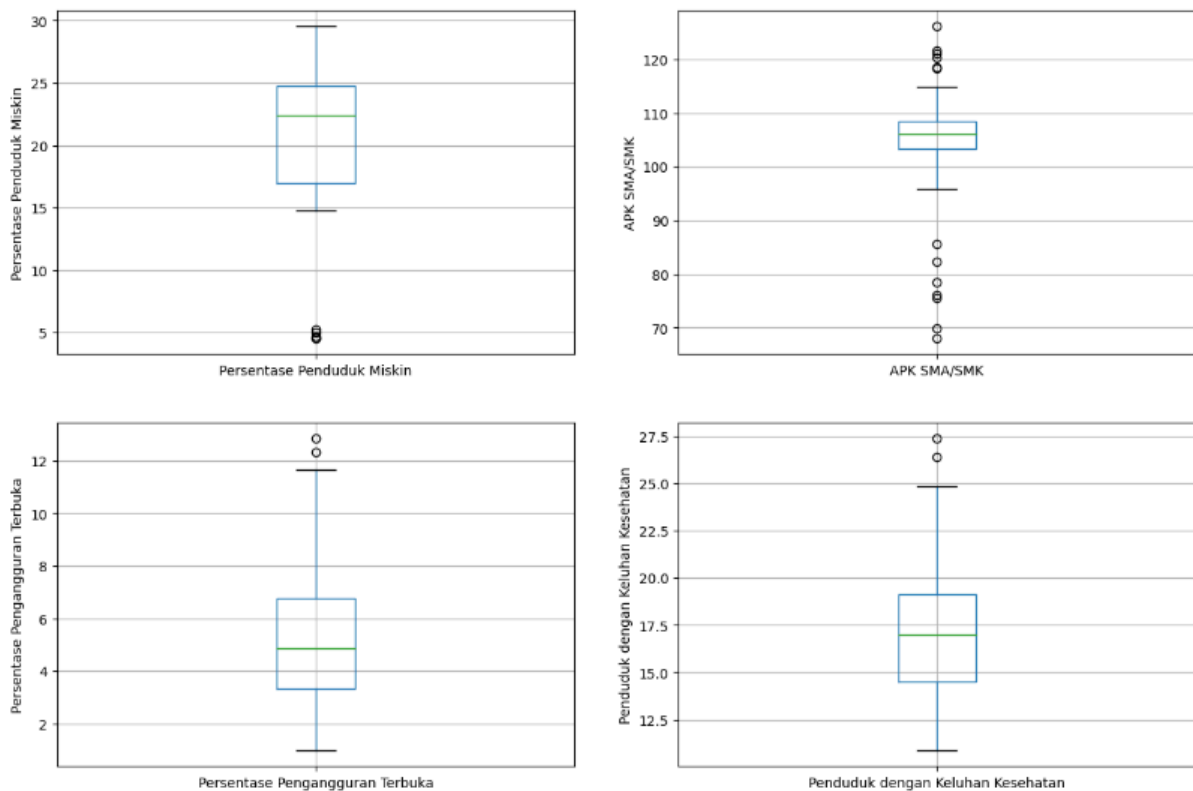
HASIL PENELITIAN DAN LUARAN YANG DICAPAI

4.1. Dataset yang digunakan

	Kabupaten/Kota	Persentase Penduduk Miskin	APK SMA/SMK	Persentase Pengangguran Terbuka	Penduduk dengan Keluhan Kesehatan	IPM
0	Maluku Tenggara Barat	27.25	112.79	4.85	16.75	0
1	Maluku Tenggara	22.75	114.61	2.58	18.34	0
2	Maluku Tengah	20.04	106.32	7.76	16.08	1
3	Buru	16.74	108.91	2.81	23.69	0
4	Kepulauan Aru	26.73	111.61	4.05	16.98	0
...
50	Seram Bagian Timur	21.08	103.92	3.31	23.13	0
51	Maluku Barat Daya	28.78	102.93	2.43	19.08	0
52	Buru Selatan	15.28	105.38	0.97	13.42	0
53	Ambon	5.25	108.04	11.65	22.67	1
54	Tual	20.68	102.76	8.98	13.50	0

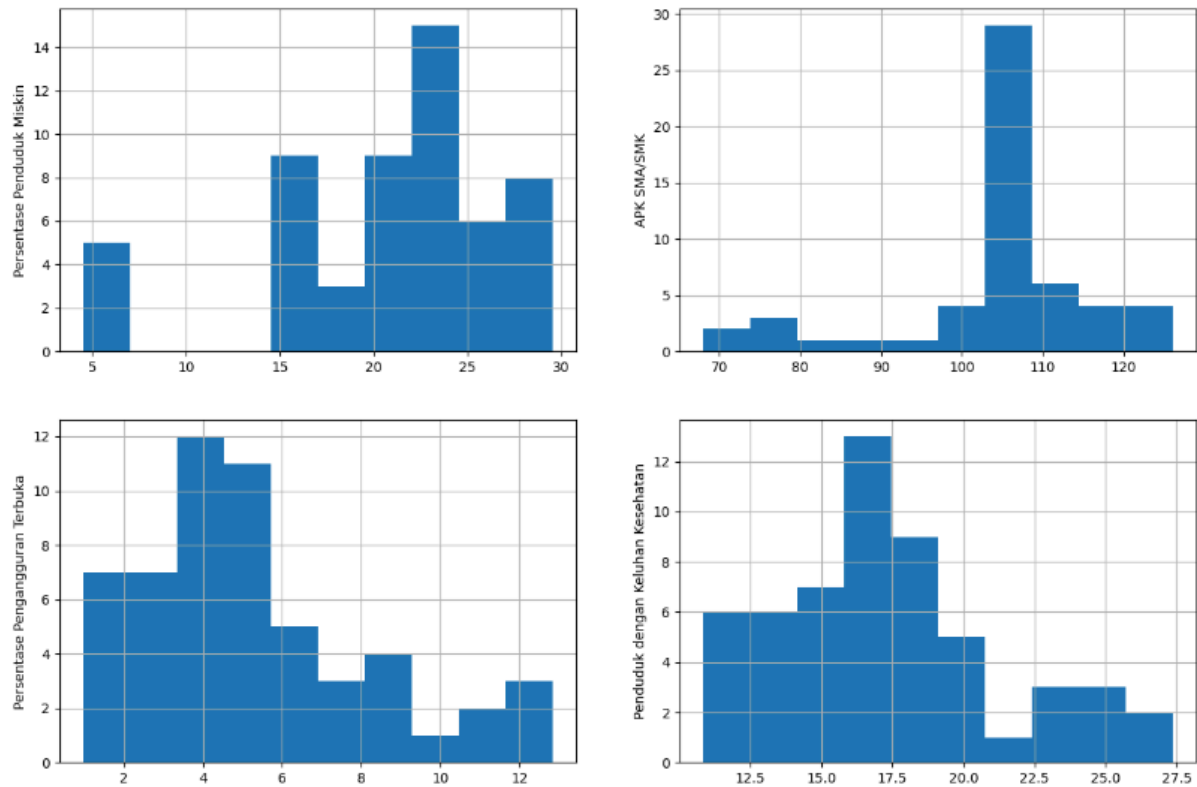
Tabel 1. Dataset

4.2. Exploratory Data Analysis (EDA)



Gambar 1. Boxplot

Plot di atas menunjukkan letak outlier untuk setiap variabel. Dimana outlier disini menandakan ketidakmerataan variabel pada tiap-tiap kabupaten di Provinsi Maluku. Pada APK.



Gambar 2. Histogram

Plot di atas menunjukkan distribusi data yang seragam.

4.3. Logistic Model

```
X = df_maluku.drop(['Kabupaten/Kota', 'IPM'], axis=1)
y = df_maluku['IPM']
y = y.astype(int)
```

```
# Split data (80/20)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

```
# Scale data (mean=0, std=1)
```

```
sc = StandardScaler()
```

```
X_train = sc.fit_transform(X_train)
```

```
X_test = sc.transform(X_test)
```

Memisahkan data variabel dan data target lalu menerapkan normalisasi untuk memastikan bahwa semua fitur dalam dataset memiliki skala yang seragam.

```

# Convert the training and testing data to torch tensors
X_train = torch.from_numpy(X_train.astype(np.float32))
y_train = torch.from_numpy(y_train.values.astype(np.float32))
X_test = torch.from_numpy(X_test.astype(np.float32))
y_test = torch.from_numpy(y_test.values.astype(np.float32))

y_train = y_train.view(-1,1)
y_test = y_test.view(-1,1)

print('y_train shape:', y_train.shape)
print('y_test shape:', y_test.shape)

# Model
class LogisticRegression(nn.Module):
    def __init__(self, n_feats):
        super().__init__()

        # Define layers
        self.lin = nn.Linear(n_feats,1)
        self.sig = nn.Sigmoid()

    def forward(self, x):
        return self.sig(self.lin(x))

model = LogisticRegression(n_feats = X.shape[1])

y_train shape: torch.Size([44, 1])
y_test shape: torch.Size([11, 1])

```



```

# Set the seed for lock initialization
torch.manual_seed(42)

# Binary cross entropy loss
loss = nn.BCELoss()

# Adam optimiser
optimiser = torch.optim.Adam(params = model.parameters(), lr=0.01)
n_iters = 190

# Loop
for epoch in range(n_iters):
    # Forward pass
    y_preds = model(X_train)
    L = loss(y_preds, y_train)

    # Backprop
    L.backward()

    # Update parameters
    optimiser.step()

    # Zero gradients
    optimiser.zero_grad()

    # Print Loss
    if epoch % 20 == 0:
        print(f'Epoch {epoch}, loss {L.item():.3f}')

```

```

Epoch 0, loss 0.822
Epoch 20, loss 0.648
Epoch 40, loss 0.531
Epoch 60, loss 0.449
Epoch 80, loss 0.389
Epoch 100, loss 0.345
Epoch 120, loss 0.311
Epoch 140, loss 0.284
Epoch 160, loss 0.262
Epoch 180, loss 0.244

```

Melakukan inisiasi model neural network dengan mendefinisikan BCELoss dan optimizer Adam dengan parameter spesifik linear. Iterasi hingga 190 untuk menghasilkan akurasi yang diinginkan.

```

with torch.no_grad():
    acc = model(X_train).round().eq(y_train).sum() / len(y_train.numpy())
    print(f'Accuracy on train set: {100*acc.item():.2f} %')

```

```
Accuracy on train set: 90.91 %
```

```

with torch.no_grad():
    acc = model(X_test).round().eq(y_test).sum() / len(y_test.numpy())
    print(f'Accuracy on test set: {100*acc.item():.2f} %')

```

```
Accuracy on test set: 90.91 %
```

```
# Get the weights and bias
weights = model.lin.weight.data.numpy()
bias = model.lin.bias.data.numpy()

# Print the equation
print("The logistic regression equation is:")
print(f' + {bias[0]}')
for i, weight in enumerate(weights[0]):
    print(f' + ({weight} * x{i+1})')
```

```
The logistic regression equation is:
+ -0.918796956539154
+ (-1.080904245376587 * x1)
+ (0.4431614875793457 * x2)
+ (1.2099543809890747 * x3)
+ (-0.21897439658641815 * x4)
```

```
# Get the weights
weights = model.lin.weight.data.numpy()

# Calculate and print the odds ratios
odds_ratios = np.exp(weights)
for i, odds_ratio in enumerate(odds_ratios[0]):
    print(f'Odds ratio for x{i+1}: {odds_ratio}')
```

```
Odds ratio for x1: 0.339288592338562
Odds ratio for x2: 1.5576238632202148
Odds ratio for x3: 3.3533318042755127
Odds ratio for x4: 0.8033422827720642
```

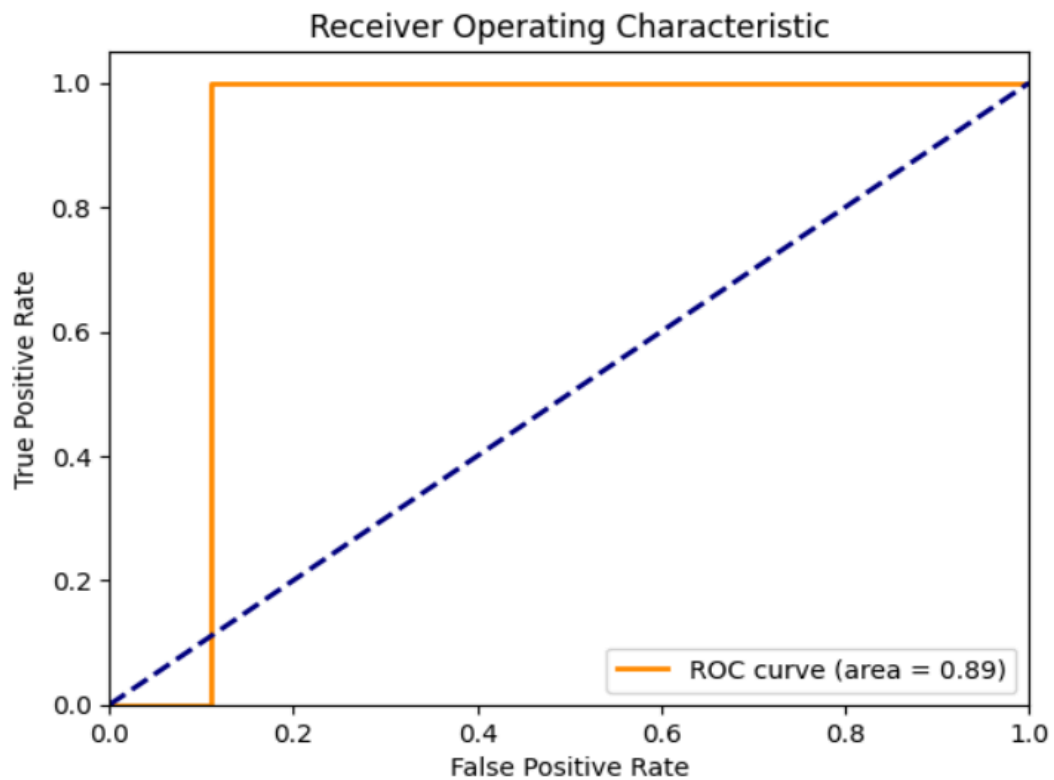
4.4. Evaluasi model

```
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve, auc

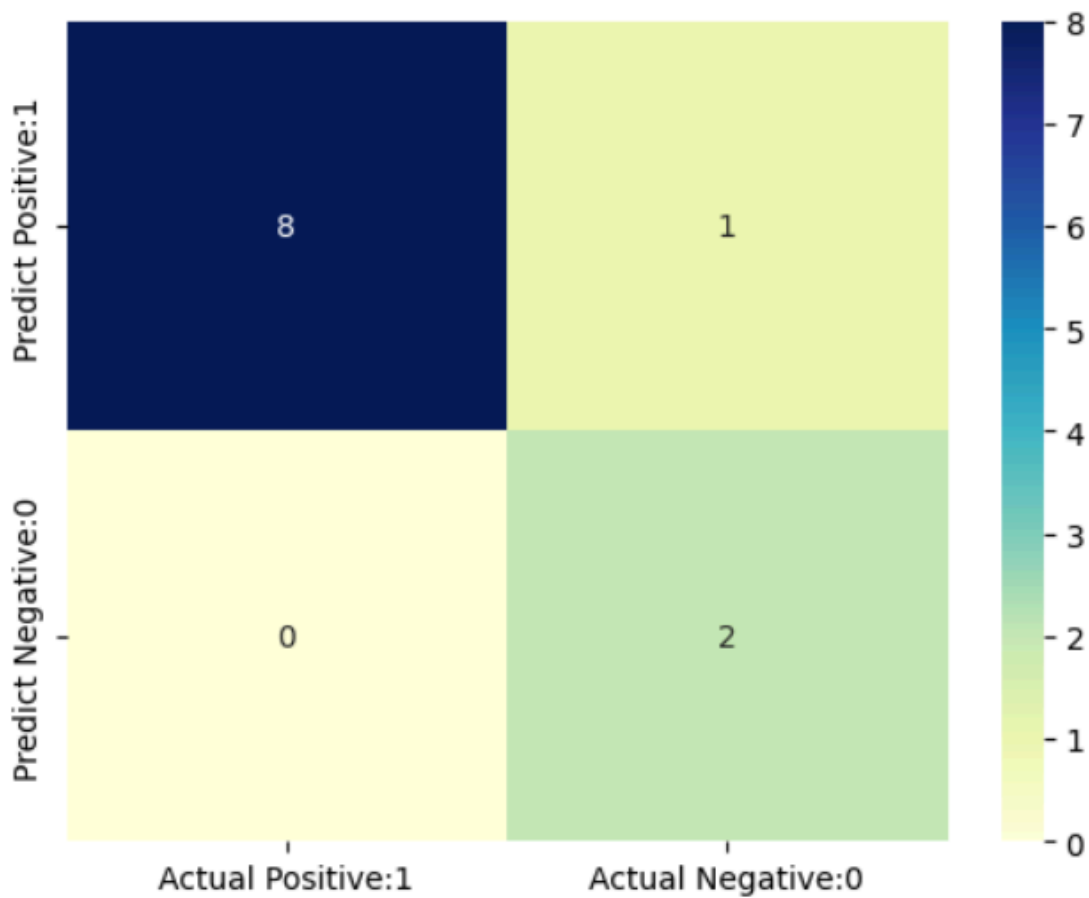
# Calculate predictions on the test set
with torch.no_grad():
    y_preds = model(X_test).numpy()

# Calculate fpr, tpr, thresholds and AUC
fpr, tpr, thresholds = roc_curve(y_test.numpy(), y_preds)
roc_auc = auc(fpr, tpr)

# Plot ROC curve
plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (area = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic')
plt.legend(loc="lower right")
plt.show()
```



Gambar 3. Plot ROC kurva



Gambar 4. Confusion matrix

```

TP = confusion_mtx_test[0,0]
TN = confusion_mtx_test[1,1]
FP = confusion_mtx_test[0,1]
FN = confusion_mtx_test[1,0]

# print classification accuracy
classification_accuracy = (TP + TN) / float(TP + TN + FP + FN)

print('Classification accuracy : {0:0.4f}'.format(classification_accuracy))

Classification accuracy : 0.9091

```

Dari hasil output klasifikasi akurasi sebesar 90%, dapat disimpulkan bahwa model yang digunakan memiliki tingkat keakuratan yang tinggi dalam memprediksi kelas target. Akurasi sebesar 90% berarti 90% dari total prediksi yang dilakukan oleh model benar. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan data dengan benar. Selain akurasi klasifikasi metrik evaluasi lainnya, seperti presisi, recall, dan f1-score untuk mendapatkan pemahaman yang lebih lengkap tentang performa model dalam mengklasifikasikan data.

```

from sklearn.metrics import classification_report

print(classification_report(y_test_true, y_test_pred))

```

	precision	recall	f1-score	support
0.0	1.00	0.89	0.94	9
1.0	0.67	1.00	0.80	2
accuracy			0.91	11
macro avg	0.83	0.94	0.87	11
weighted avg	0.94	0.91	0.92	11

```

# print classification error

classification_error = (FP + FN) / float(TP + TN + FP + FN)

print('Classification error : {0:0.4f}'.format(classification_error))

Classification error : 0.0909

```

```

# Prediksi from data testing
with torch.no_grad():
    y_test_pred = model(X_test).round().numpy()

# Save prediction
test_results = pd.DataFrame({'Actual': y_test.numpy().flatten(), 'Predicted': y_test_pred})
print(test_results)

```

	Actual	Predicted
0	0.0	0.0
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.0	0.0
5	1.0	1.0
6	0.0	0.0
7	0.0	1.0
8	1.0	1.0
9	0.0	0.0
10	0.0	0.0

BAB 5

KESIMPULAN DAN SARAN

5.1. Interpretasi Hasil

Persamaan regresi logit biner berdasarkan model di atas dituliskan sebagai berikut.

$$y = -1 + (-1 * X1) + (0.4 * X2) + (1.2 * X3) - (0.2 * X4)$$

Di sini (X1), (X2), (X3), dan (X4) adalah variabel input, dan (y) adalah variabel output. Ini adalah bentuk umum dari persamaan linear untuk model regresi logistik. Karena nilai koefisien regresi logit sulit untuk dijelaskan secara langsung dari model yang diperoleh. Untuk itu, dalam penjelasan dari model regresi logistik ini dilakukan dengan antilog dari B atau pangkat eksponensial dari koefisien logit yang biasa disebut sebagai odds ratio. Odds ratio untuk X1 sebesar 0.339 dengan koefisien negatif, artinya ketika persentase penduduk miskin meningkat sebesar 1% maka rasio indeks pembangunan manusia akan turun sebesar 0.339 dengan asumsi variabel lainnya tetap.

Odds ratio untuk X2 sebesar 1.557 artinya ketika pengeluaran perkapita meningkat 1 ribu rupiah/orang/tahun maka rasio Indeks Pembangunan Manusia (IPM) naik sebesar sebesar 1.557 dengan asumsi variabel lainnya tetap. Odds ratio untuk X3 sebesar 3.353 artinya ketika persentase pengangguran terbuka naik 1% maka Indeks Pembangunan Manusia naik sebesar sebesar 3.353 dengan asumsi variabel lainnya tetap. Odds ratio untuk X4 sebesar 0.803 dengan koefisien negatif artinya ketika penduduk dengan keluhan kesehatan naik 1% maka Indeks Pembangunan Manusia turun sebesar sebesar 0.803 dengan asumsi variabel lainnya tetap.

5.2. Kesimpulan

Berdasarkan hasil yang diperoleh, model regresi logistik biner telah menunjukkan kinerja yang sangat baik dengan akurasi 90% baik pada data latihan maupun data pengujian. Konsistensi ini menunjukkan bahwa model telah belajar dengan baik dari data latihan dan mampu menggeneralisasi hasilnya ke data yang belum pernah dilihat sebelumnya (data pengujian) dengan tingkat akurasi yang sama. Selain itu, model ini tidak menunjukkan tanda-tanda *overfitting* dan *underfitting*, karena performa model pada data latihan dan pengujian hampir sama. *Overfitting* biasanya ditandai dengan performa yang sangat baik pada data latihan tetapi buruk pada data pengujian, sedangkan *underfitting* ditandai dengan performa yang buruk pada kedua jenis data tersebut.

Secara keseluruhan, model regresi logistik biner ini telah menunjukkan kinerja yang sangat baik dan dapat digunakan sebagai alat yang efektif untuk memprediksi IPM

sedang atau tinggi. ROC AUC model sebesar 0.89 berarti, model klasifikasi logit biner bekerja dengan sangat baik memprediksi IPM sedang atau tinggi.

5.3. Saran

Pada penelitian selanjutnya dapat pula menambahkan variabel yang diduga mempengaruhi indeks pembangunan manusia seperti pertumbuhan ekonomi, tingkat inflasi serta tingkat pengangguran atau tingkat angkatan kerja. Selain itu, dapat juga dilakukan klasifikasi dengan menggunakan regresi logistik ordinal dengan indeks pembangunan manusia dengan kategori rendah, sedang dan tinggi agar dapat diperoleh hasil yang lebih sesuai. Pemerintah Provinsi Maluku harus memperhatikan proses peningkatan dan pemerataan pembangunan di setiap kabupaten/kota di Provinsi Maluku untuk setiap sektor, baik dari segi kesehatan, pendidikan maupun ekonomi.

DAFTAR PUSTAKA

- Anisa, L., & Rifai, N. A. K. (2022). Analisis Regresi Logistik Biner dengan Metode PMLE pada Penyakit Covid-19. *Jurnal Riset Statistika*, 2(2), 129–136.
- Astuti, W. I., Ratnasari, V., & Wibowo, W. (2017). Analisis Faktor yang Berpengaruh Terhadap Tingkat Pengangguran Terbuka di Provinsi Jawa Timur Menggunakan Regresi Data Panel. *Jurnal Sains Dan Seni ITS*, 6(1). <https://doi.org/10.12962/j23373520.v6i1.22977>
- Ferezagia, D. V. (2018). Analisis Tingkat Kemiskinan di Indonesia. *Jurnal Sosial Humaniora Terapan*, 1(1), 1–7.
- Manurung, E. N., & Hutabarat, F. (2021). Pengaruh Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Pengeluaran per Kapita Terhadap Indeks Pembangunan Manusia. *Jurnal Ilmiah Akuntansi Manajemen*, 4(2), 121–129. <https://doi.org/10.35326/jiam.v4i2.1718>
- Mukhaiyar, U., Rontos, F., Handoko, K., & Kardiyaniti, S. (2022). Analisis Faktor-Faktor yang Memengaruhi Angka Partisipasi Kasar SMA/Sederajat di Indonesia Menggunakan Regresi Ridge. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 10(2), 222–234. <https://doi.org/10.34312/euler.v10i2.15903>
- Pratowo, N. I. (2012). ANALISIS FAKTOR-FAKTOR YANG BERPENGARUH TERHADAP INDEKS PEMBANGUNAN MANUSIA . *Jurnal Studi Ekonomi Indonesia*, 1(1), 15–31.
- Retnawati, R. (2017). Pengantar Analisis Regresi dan Korelasi. In *Workshop Teknik Analisis Data Fakultas Ekonomi Dan Bisnis IAIN Batusangkar Di Rocky Hotel Bukittinggi*, 25.
- Rofiqo, N., Windarto, A. P., & Hartama, D. (2018). PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN DENGAN DATAMINING K-MEANS. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1), 216–223.
- Susetyoko, R., Yuwono, W., & Purwantini, E. (2022). Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner. *JIP (Jurnal Informatika Polinema)*, 8(4), 31–40.