

## LAMPIRAN

### Lampiran 1. Dataset Penelitian (Data Ketenagakerjaan)

Dataset yang digunakan dalam penelitian ini diperoleh dari Dinas Tenaga Kerja Kabupaten Labuhanbatu yang berisi 50 data pekerja. Variabel yang digunakan meliputi:

1. Nama
2. Tanggal Lahir
3. Usia
4. Jenis Kelamin
5. Pendidikan Terakhir
6. Pengalaman Kerja
7. Status Pernikahan
8. Kecamatan
9. Status Pekerjaan

Contoh data:

<b>Nama</b>	<b>Usia</b>	<b>Pengalaman</b>	<b>Status Pernikahan</b>	<b>Kecamatan</b>
Budi Nasution	47	12	Menikah	Rantau Utara
Sari Siregar	48	7	Menikah	Bilah Hulu
Khairunnisa Aisyah	26	2	Menikah	Bilah Barat

### Lampiran 2. Hasil Analisis Missing Value

Berdasarkan hasil pemeriksaan dataset:

<b>Kolom</b>	<b>Missing Value</b>
Nama	0
Usia	0
Jenis Kelamin	0
Pendidikan	0
Pengalaman Kerja	0
Status Pekerjaan	0

### **Kesimpulan:**

Dataset tidak memiliki missing value sehingga tidak diperlukan proses imputasi.

### **Lampiran 3. Konversi Data Kategorikal ke Numerik**

Proses transformasi data:

1. Jenis Kelamin:
  - a. Laki - laki = 0
  - b. Perempuan = 1
2. Pendidikan:
  - a. SMP = 1
  - b. SMA = 2
  - c. Diploma = 3
  - d. Sarjana = 4
3. Status Pekerjaan:
  - a. Bekerja = 1
  - b. Tidak Bekerja = 0

Contoh hasil konversi:

<b>JK</b>	<b>Pendidikan</b>	<b>Status</b>
0	2	1
1	3	0
1	4	1

### **Lampiran 4. Normalisasi Data**

Metode yang digunakan: **Min-Max Normalization**

Contoh hasil:

<b>Usia (Norm)</b>	<b>Pengalaman (Norm)</b>
0.176	0.142
1.000	0.357
0.411	0.285

### **Tujuan:**

Menyamakan skala data agar tidak terjadi bias pada model.

### Lampiran 5. Pembagian Data Training dan Testing

- a. Data Training: 80% (40 data)
- b. Data Testing: 20% (10 data)

Contoh data training:

<b>Nama</b>	<b>Usia</b>	<b>Pengalaman</b>
Budi Nasution	47	12
Haura Zalfa	25	3

Contoh data testing:

<b>Nama</b>	<b>Usia</b>	<b>Pengalaman</b>
Budi Hutabarat	42	10
Khalisa Amira	16	0

### Lampiran 6. Perhitungan Gini Impurity

Contoh perhitungan:

<b>Kelompok</b>	<b>Jumlah</b>	<b>Gini</b>
Usia $\leq$ 30	8	0.46875
Usia $>$ 30	2	0.5
Total	10	0.475

### Lampiran 7. Pemilihan Fitur

<b>Fitur</b>	<b>Gini</b>
Usia $\leq$ 30	0.424
Usia $>$ 30	0.408
Pengalaman $\leq$ 5	0.437
Pengalaman $>$ 5	0.375

### Kesimpulan:

Fitur terbaik adalah **Pengalaman Kerja** karena memiliki Gini terendah.

## Lampiran 8. Random Forest (Bootstrap Sampling)

Pohon	Gini
Tree 1	0.4712
Tree 2	0.4200
Tree 3	0.3200
Tree 4	0.3432
Tree 5	0.4032

## Lampiran 9. Script Python (Implementasi Model)

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.metrics import classification_report

# Load data
data = pd.read_csv('data_ketenagakerjaan.csv')

X = data.drop('Status', axis=1)
y = data['Status']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Random Forest
rf = RandomForestClassifier()
rf.fit(X_train, y_train)
rf_pred = rf.predict(X_test)

# SVM
svm = SVC()
svm.fit(X_train, y_train)
svm_pred = svm.predict(X_test)

# Evaluasi
print("Random Forest")
print(classification_report(y_test, rf_pred))

print("SVM")
print(classification_report(y_test, svm_pred))
```

## **Lampiran 10. Hasil Evaluasi Model**

Metrik yang digunakan:

1. Accuracy
2. Precision
3. Recall
4. F1-Score

### **Kesimpulan:**

- a. Random Forest → lebih stabil & akurat
- b. SVM → baik pada data terstruktur

## **Lampiran 11. Dokumentasi Penelitian**

1. Lokasi: Dinas Tenaga Kerja Kabupaten Labuhanbatu
2. Tools: Python, Scikit-learn, Google Colab
3. Dataset: Data Ketenagakerjaan