Tugas 2: - Tugas Praktikum Mandiri

ROHMATUL HIDAYAT - 0110224015

Teknik Informatika, STT Terpadu Nurul Fikri, Depok

*E-mail: 0110224015@student.nurulfikri.ac.id –

1. Membaca Dataset

```
# Membaca Lokasi Data
df = pd.read_csv('../Data/day.csv')
```

Penjelasan:

- Membaca file day.csv dari folder Data
- Menggunakan pandas untuk membaca data CSV ke dalam DataFrame

Output:

```
~~~ DATASET AWAL ~~~
Jumlah total data: 731
5 baris pertama:
   instant      dteday  season  yr  mnth  holiday  weekday  workingday  \
0        1  2011-01-01       1   0     1        0        6           0
1        2  2011-01-02       1   0     1        0        0           0
2        3  2011-01-03       1   0     1        0        1           1
3        4  2011-01-04       1   0     1        0        2           1
4        5  2011-01-05       1   0     1        0        3           1
```

2. Menghitung Pembagian Data

```
# Mengitung jumlah untuk setiap bagian
totalData = len(df)
training = int(totalData * 0.8)  # 80% untuk training
testing = totalData - training  # 20% untuk testing
validation = int(training * 0.1)    # 10% dari training untuk validation
```

Penjelasan:

- totalData: Menghitung total baris data (731)
- training: 80% dari total data = 584 baris
- testing: Sisa 20% dari total data = 147 baris
- validation: 10% dari data training = 58 baris

3. Acak Data terlebih dahulu sebelum membagi data

```
# Acak data
df_shuffled = df.sample(frac=1, random_state=42).reset_index(drop=True)
```

Penjelasan:

- sample(frac=1): Mengambil 100% data secara acak
- random_state=42: Untuk hasil yang konsisten/reproducible
- reset_index(drop=True): Mereset index setelah pengacakan

4. Membagi Data yang sudah diacak

```
# Bagi data
trainingData = df_shuffled[:training - validation]
validationData = df_shuffled[training - validation:training]
testingData = df_shuffled[training:]

print("---------------------------------------- HASIL PEMBAGIAN ----------------------------------------")
print(f"Data Training: {len(trainingData)} baris")
print(f"Data Validation: {len(validationData)} baris")
print(f"Data Testing: {len(testingData)} baris")
print()

print("---------------------------------------- DATA TRAINING ----------------------------------------")
print(trainingData.head())
print()

print("---------------------------------------- DATA VALIDATION ----------------------------------------")
print(validationData.head())
print()

print("---------------------------------------- DATA TESTING ----------------------------------------")
print(testingData.head())
```

0.0s

Penjelasan:

- **Training**: 584 - 58 = 526 baris (indeks 0-525)
- **Validation**: 58 baris (indeks 526-583)
- **Testing**: 147 baris (indeks 584-730)

OUTPUT HASIL PEMBAGIAN:

```
---------------------------------------- HASIL PEMBAGIAN ----------------------------------------
Data Training: 526 baris
Data Validation: 58 baris
Data Testing: 147 baris
```

OUTPUT DATA TRAINING:

```
---------------------------------------- DATA TRAINING ----------------------------------------
   instant      dteday  season  yr  mnth  holiday  weekday  workingday  \
0      704  2012-12-04       4   1    12        0        2           1
1       34  2011-02-03       1   0     2        0        4           1
2      301  2011-10-28       4   0    10        0        5           1
3      457  2012-04-01       2   1     4        0        0           0
4      634  2012-09-25       4   1     9        0        2           1

   weathersit      temp     atemp       hum  windspeed  casual  registered  \
0           1  0.475833  0.469054  0.733750   0.174129     551        6055
1           1  0.186957  0.177878  0.437826   0.277752      61        1489
2           2  0.330833  0.318812  0.585833   0.229479     456        3291
3           2  0.425833  0.417287  0.676250   0.172267    2347        3694
4           1  0.550000  0.544179  0.570000   0.236321     845        6693

    cnt
0  6606
1  1550
2  3747
3  6041
4  7538
```

OUTPUT DATA VALIDATION:

```
---------------------------------------- DATA VALIDATION ----------------------------------------
     instant      dteday  season  yr  mnth  holiday  weekday  workingday  \
526      677  2012-11-07       4   1    11        0        3           1
527      549  2012-07-02       3   1     7        0        1           1
528      349  2011-12-15       4   0    12        0        4           1
529      464  2012-04-08       2   1     4        0        0           0
530      326  2011-11-22       4   0    11        0        2           1

     weathersit      temp     atemp       hum  windspeed  casual  registered  \
526           2  0.295833  0.274621  0.547500   0.304108     326        4709
527           1  0.781667  0.702038  0.447083   0.195267     904        5323
528           2  0.422500  0.412237  0.634167   0.268042     181        3528
529           1  0.500000  0.492425  0.275833   0.232596    2230        2939
530           3  0.416667  0.421696  0.962500   0.118792      69        1538

      cnt
526  5035
527  6227
528  3709
529  5169
530  1607
```

OUTPUT DATA TESTING

```
---------------------------------------- DATA TESTING ----------------------------------------
     instant      dteday  season  yr  mnth  holiday  weekday  workingday  \
584      658  2012-10-19       4   1    10        0        5           1
585      624  2012-09-15       3   1     9        0        6           0
586      617  2012-09-08       3   1     9        0        6           0
587      557  2012-07-10       3   1     7        0        2           1
588      578  2012-07-31       3   1     7        0        2           1

     weathersit      temp     atemp       hum  windspeed  casual  registered  \
584           2  0.563333  0.537896  0.815000   0.134954     753        4671
585           1  0.608333  0.585867  0.501667   0.247521    3160        5554
586           2  0.659167  0.611121  0.799167   0.281104    1557        4419
587           2  0.720833  0.664796  0.667500   0.151737     954        5336
588           1  0.713333  0.662896  0.704167   0.165425     968        6248

      cnt
584  5424
585  8714
586  5976
587  6290
588  7216
```

KESIMPULAN HASIL IMPLEMENTASI ALGORITMA

1. **Proporsi Optimal Terpenuhi**

- Training Set: 526 baris (72% dari total)
- Validation Set: 58 baris (8% dari total)
- Test Set: 147 baris (20% dari total)

2. **Kualitas Pembagian Data**

- **Data teracak sempurna** - mencegah bias temporal
- **Distribusi merata** - semua subset mewakili populasi
- **Konsistensi terjaga** dengan random_state=42