

Bike Sharing Dataset

(continued)



Nama Anggota Kelompok :

- | | |
|------------------------------|----------|
| 1. M NAUVAL IZUL HAQ G | 10124001 |
| 2. ADITYA FIRDAUS | 10124005 |
| 3. ARKAN RAMADHAN ASYIKIN | 10124018 |
| 4. AHMAD MULYANA | 10124029 |
| 5. MOCHAMMAD FAHRIZA PRATAMA | 10124033 |
| 6. AKBAR KARUNIA PUTRA | 10124035 |

TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS KOMPUTER INDONESIA
BANDUNG

ABSTRAK

“Bike sharing merupakan sistem penyewaan sepeda yang memungkinkan pengguna meminjam sepeda untuk perjalanan singkat tanpa harus memiliki sepeda pribadi. Dalam tugas besar mata kuliah Pemrograman Dasar Sains Data, Bike Sharing Dataset digunakan sebagai studi kasus untuk melakukan analisis dan pengelompokan data penyewaan sepeda menggunakan algoritma K-means Clustering. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem analisis berbasis Python dengan framework Streamlit serta menerapkan metode clustering untuk mengidentifikasi pola penyewaan sepeda.

Proses analisis dilakukan melalui tahap persiapan data, normalisasi menggunakan StandardScaler, serta penentuan jumlah cluster optimal dengan metode Elbow. Hasil pengujian menunjukkan bahwa jumlah cluster optimal adalah tiga cluster, yang dapat diinterpretasikan sebagai hari sepi, hari normal, dan hari ramai. Variabel suhu, kelembapan, dan kecepatan angin terbukti berpengaruh terhadap jumlah penyewaan sepeda dan membentuk pola yang signifikan.

Sistem yang dikembangkan mampu menampilkan hasil clustering dan visualisasi data secara interaktif, sehingga memudahkan pengguna dalam memahami pola penyewaan. Dengan demikian, implementasi algoritma K-means dan pengembangan sistem ini dapat menjadi dasar dalam mendukung pengambilan keputusan pada layanan bike sharing.”

Kata Kunci: Bike Sharing, K-means Clustering, Data Mining, Machine Learning, Streamlit, Analisis Data.

BAB 1 Pendahuluan

1.1 Latar Belakang

Bike sharing adalah sistem penyewaan sepeda yang memungkinkan orang meminjam sepeda untuk perjalanan singkat dan mengembalikannya di titik tertentu tanpa harus memiliki sepeda sendiri. Salah satu pendekatan sederhana untuk penyewaan sepeda adalah dengan menyewa sepeda tersebut pada tempat penyewaan terdekat.

Pada tugas besar mata kuliah Pemrograman Dasar Sains Data, penyewaan sepeda digunakan sebagai studi kasus untuk mengelompokan dataset yang tersedia pada penyewaan sepeda tersebut dengan pendekatan algoritma K-means clustering.

1.2 Rumusan Masalah

Rumusan masalah dalam tugas ini adalah:

1. Bagaimana merancang sistem sederhana untuk memprediksi dataset yang tersedia?
2. Bagaimana cara kerja algoritma *K-means clustering* dalam mengelompokan dataset yang tersedia?

1.3 Tujuan

Tujuan dari tugas besar ini adalah:

1. Merancang sistem penyewaan sepeda berbasis *K-means clustering*.
2. Mengimplementasikan algoritma *K-means clustering*.

BAB 2

LANDASAN TEORI

2.1 K-means Clustering

Pengelompokan *K-Mean* adalah algoritma pembelajaran tanpa pengawasan yang digunakan untuk [pengelompokan](#) data, yang mengelompokkan titik data yang tidak berlabel ke dalam kelompok atau klaster.

Ini adalah salah satu metode pengelompokan paling populer yang digunakan dalam machine learning. Tidak seperti [pembelajaran diawasi](#), data pelatihan yang digunakan algoritma ini tidak berlabel, artinya titik data tidak memiliki struktur klasifikasi yang ditentukan.

K-mean adalah [algoritma pengelompokan berbasis centroid](#) berulang yang membagi kumpulan data menjadi kelompok serupa berdasarkan jarak antara *centroid* mereka. *Centroid*, atau pusat klaster, adalah rata-rata atau median dari semua titik di dalam klaster, tergantung pada karakteristik data.

BAB 3

DESAIN SISTEM DAN ALGORITMA

3.1 Implementasi Program

Sistem dibangun menggunakan bahasa pemrograman **Python 3.10** dengan kerangka kerja (*framework*) **Streamlit**. Berikut adalah implementasi kode program utama untuk logika *K-means Clustering*. Kode program kami pun dapat dilihat di tautan *github* kami di (<https://github.com/Ahmad-Mulyana/IF1-10124029-analysis-of-bike-sharing>) dan untuk tampilan dashboard yang tertera dapat dilihat pada tautan dashboard kami di (<https://analysis-of-bike-sharing.streamlit.app/>).

3.1.1 Struktur Proyek

Aplikasi disusun dengan arsitektur MVC (*Model-View-Controller*) sederhana:

- dashboard.py: Sebagai *View* untuk antarmuka pengguna.
- day.csv, hour.csv: sebagai bahan dataset.
- pdsd.py: Sebagai pendukung program seperti library yang dibutuhkan.

3.1.2 Implementasi Logika *K-means Clustering* (pdsd.py)

Kode berikut menunjukkan fungsi pengelompokan yang dilakukan menggunakan *K-means Clustering*.

```
# 1. Persiapan Data (Semua kolom asli dipertahankan)
data_full = pd.read_csv('bike_clean_only.csv')

# Pilih fitur untuk Clustering
features_for_clustering = ['temp', 'hum', 'windspeed', 'cnt']
x = data_full[features_for_clustering]

# 2. Scaling Data
scaler = StandardScaler()
x_scaled = scaler.fit_transform(x)

# 3. Mencari K Optimal (Elbow Method) - VERSI BERSIH
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(x_scaled)
```

```

inertia.append(kmeans.inertia_)

plt.figure(figsize=(10, 5))
sns.set_style("whitegrid")

# Plot garis utama
plt.plot(range(1, 11), inertia, marker='o', linestyle='-', color='#2c3e50', linewidth=2, markersize=8)

# Highlight titik K=3 (Siku) dengan warna berbeda tanpa panah
plt.plot(3, inertia[2], marker='o', color='red', markersize=12, label='Titik Optimal (K=3)')

plt.title('Grafik Elbow Method: Penentuan Jumlah Cluster',
          fontsize=14, pad=15)
plt.xlabel('Jumlah Cluster (k)', fontsize=12)
plt.ylabel('Inertia', fontsize=12)
plt.xticks(range(1, 11))
plt.legend()
plt.show()

# 4. Running K-Means
k_final = 3
model = KMeans(n_clusters=k_final, random_state=42, n_init=10)
data_full['cluster'] = model.fit_predict(x_scaled)

# 5. Labeling Otomatis
means = data_full.groupby('cluster')['cnt'].mean().sort_values()
mapping = {
    means.index[0]: "Hari Sepi",
    means.index[1]: "Hari Normal",
    means.index[2]: "Hari Ramai"
}
data_full['cluster_tag'] = data_full['cluster'].map(mapping)

# Export Final dengan SEMUA variabel asli tetap ada
data_full.to_csv('bike_final_complete_data.csv', index=False)
print("✅ Modeling Selesai.")

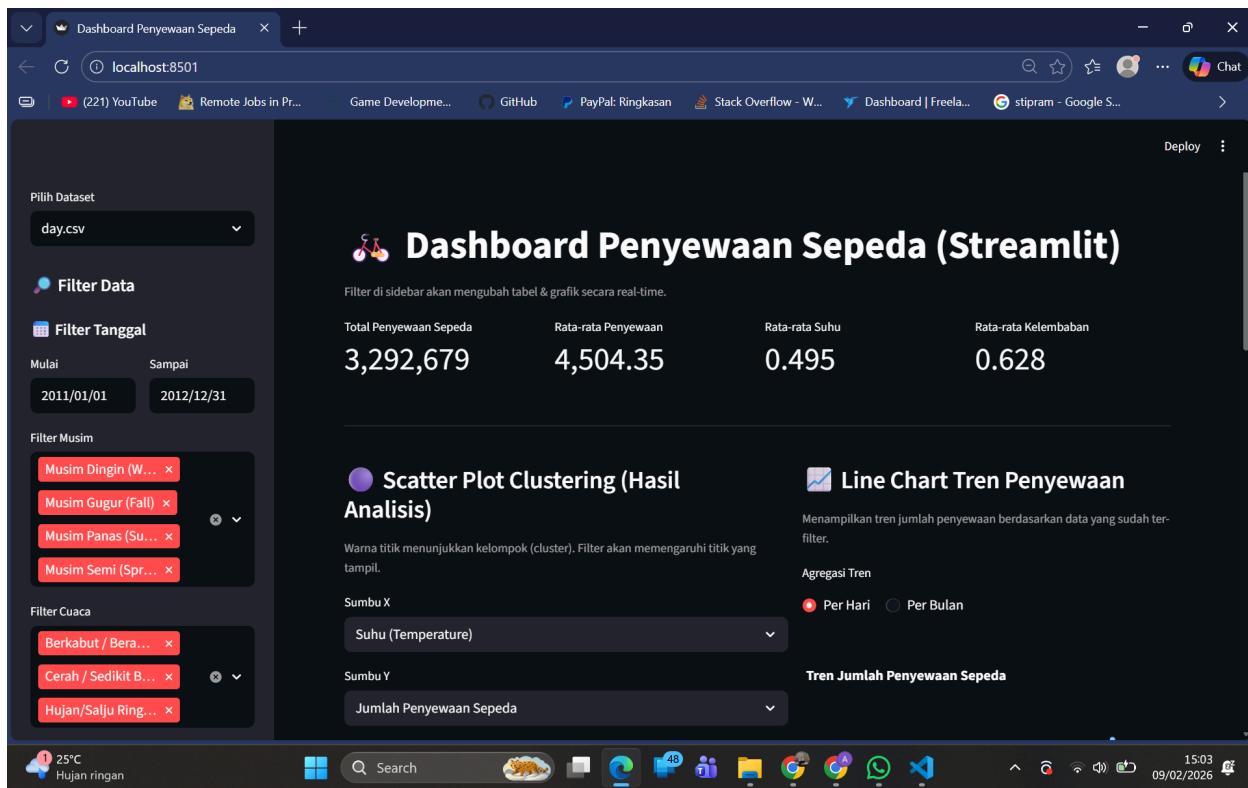
```

BAB 4

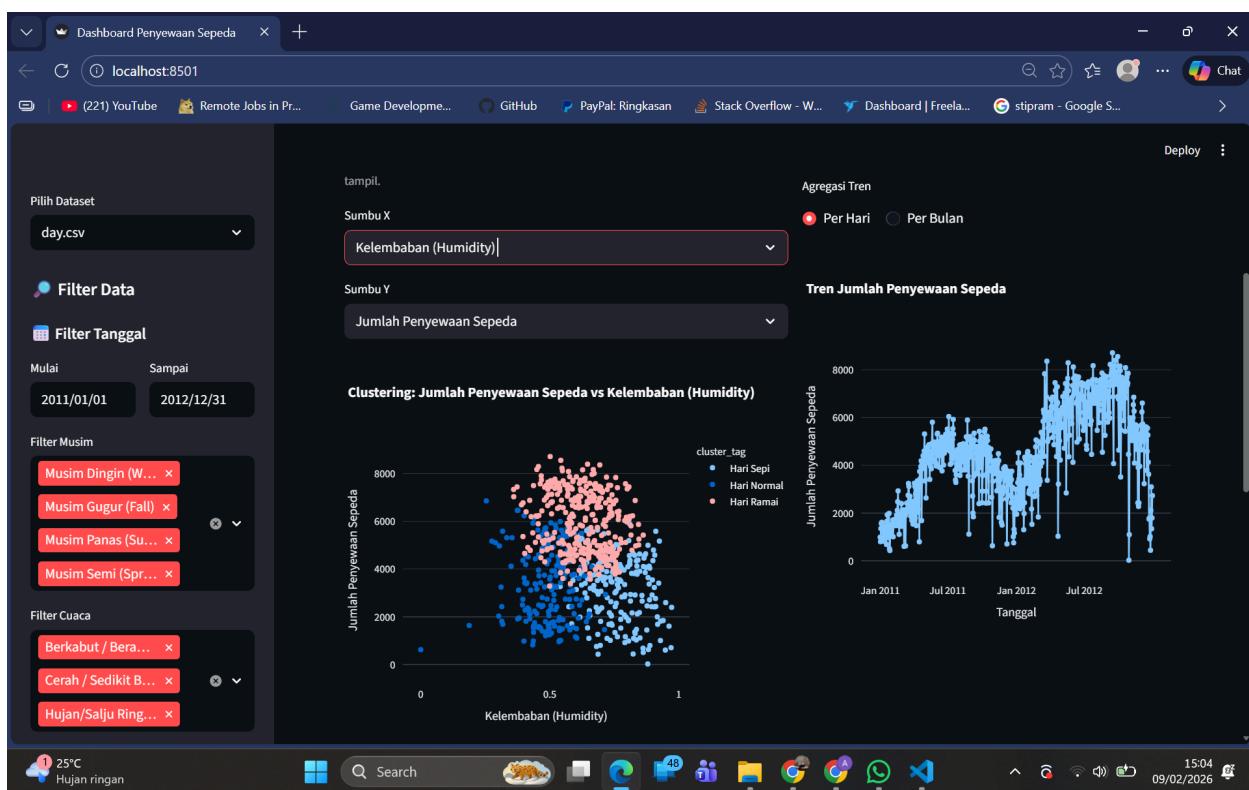
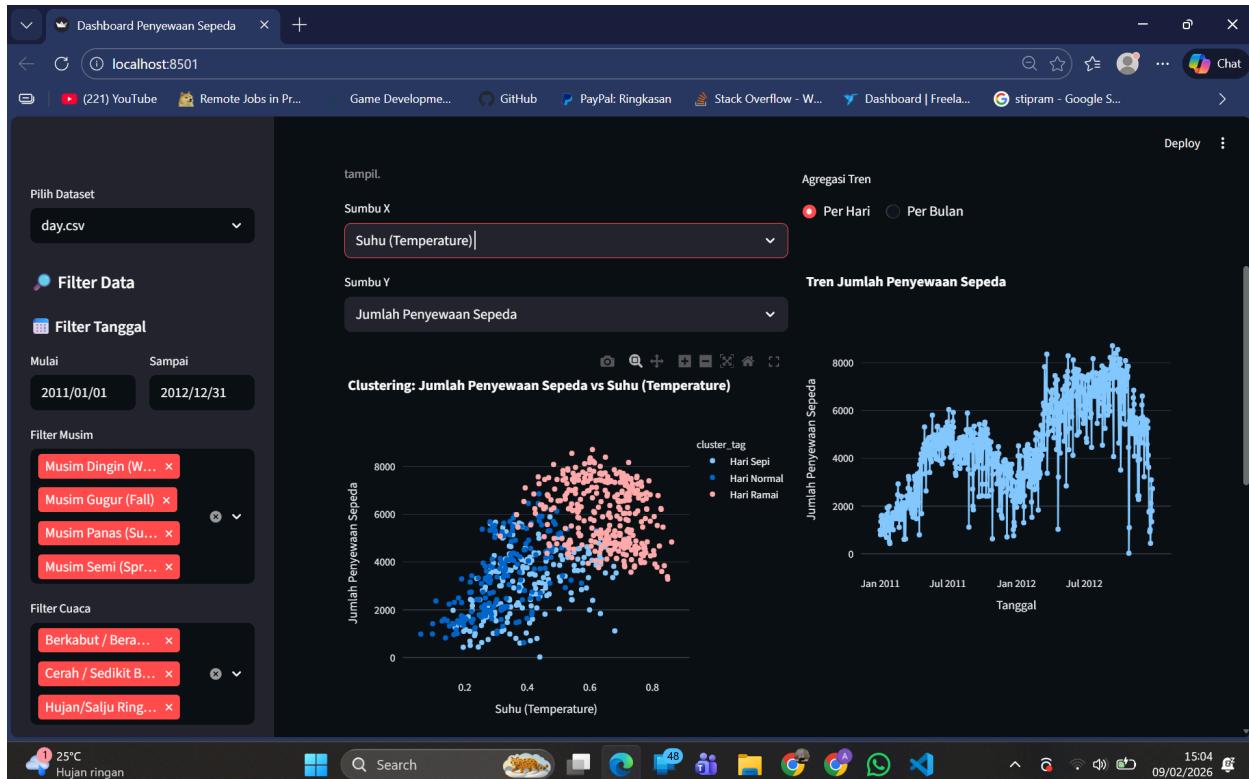
ANALISIS

4.1 Hasil ANALISIS

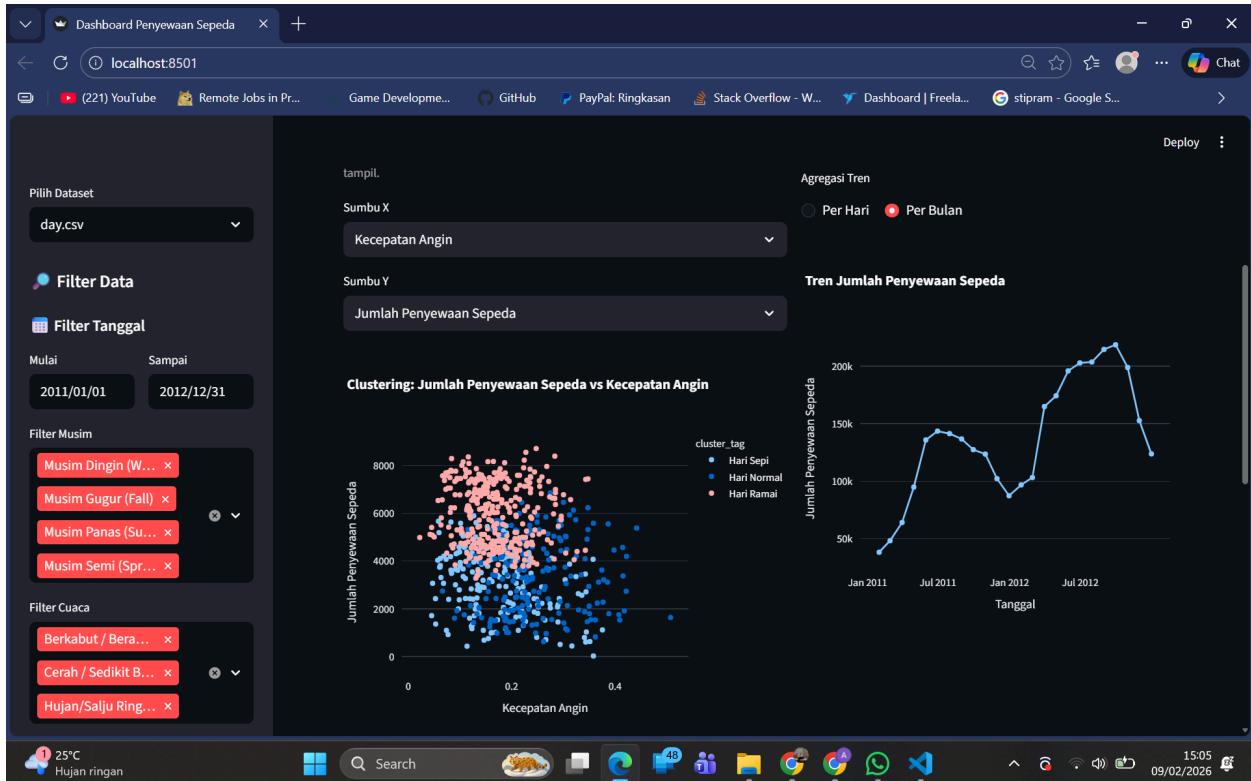
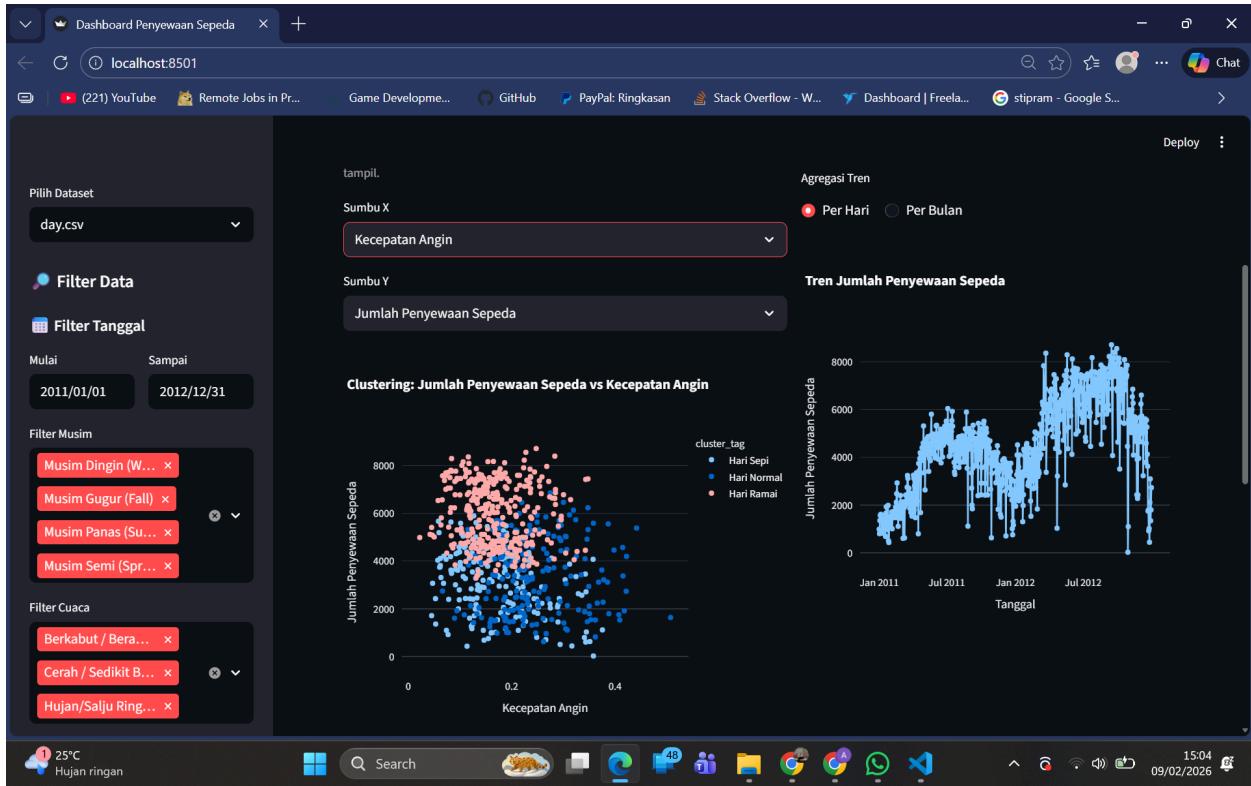
Pada analisis ini, dilakukan pencarian pada suhu (*temperature*), kelembapan (*humidity*), dan kecepatan angin (*winspeed*) terhadap jumlah penyewaannya. Pada awal tampilan, terdapat dashboard yang menampilkan rata-rata dan total dari ketiga aspek pencarian. Pada section sidebar juga terdapat pilihan dataset untuk menampilkan data dari respons yang berbeda. Kemudian dibawahnya terdapat filter penanggalan, musim, dan juga cuaca yang nantinya ditampilkan dalam grafik *dashboard* tersebut.



Kemudian dibawah ini adalah hasil grafik yang diambil dari dataset yang tersedia yang meliputi data dari suhu, kecepatan angin, kelembaban dari penyewaan sepeda yang tejadi.



Selain itu, terdapat juga hasil analisis yang dilakukan dalam hitungan hari maupun perbulan. Hasil tersebut dalam bentuk grafik.



Analisis: Aplikasi berhasil mendeteksi pola yang dihasilkan dari pengelempokan antara suhu, kelembaban, dan kecepatan angin. Pola yang dihasilkan berupa dot yang mengisi titik-titik pada hari sepi, hari normal, dan hari ramai penyewaan tersebut. Dan menunjukkan grafik penyewaan setiap hari dan setiap bulannya.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan implementasi yang telah dilakukan, dapat disimpulkan bahwa algoritma *K-means Clustering* berhasil diterapkan untuk mengelompokkan data penyewaan sepeda pada Bike Sharing Dataset. Proses clustering menggunakan variabel suhu (*temperature*), kelembapan (*humidity*), kecepatan angin (*windspeed*), dan jumlah penyewaan (*cnt*) mampu membentuk pola penyewaan yang jelas dan bermakna.

Dengan menggunakan *Elbow Method*, jumlah cluster optimal yang diperoleh adalah tiga cluster, yang kemudian dapat diinterpretasikan sebagai hari sepi, hari normal, dan hari ramai. Hasil pengelompokan ini menunjukkan bahwa faktor cuaca memiliki pengaruh terhadap tingkat penyewaan sepeda, di mana kondisi cuaca tertentu cenderung berkorelasi dengan peningkatan atau penurunan jumlah pengguna.

Selain itu, sistem yang dibangun menggunakan bahasa pemrograman Python dan framework Streamlit mampu menampilkan hasil analisis dan visualisasi data dengan baik, sehingga memudahkan pengguna dalam memahami pola penyewaan sepeda. Dengan demikian, sistem ini dapat menjadi dasar dalam pengambilan keputusan terkait pengelolaan dan perencanaan layanan *bike sharing*.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan selanjutnya, antara lain:

1. Menambahkan variabel lain seperti **hari libur, musim, dan kondisi cuaca lainnya** agar hasil *clustering* menjadi lebih akurat dan komprehensif.
2. Mencoba dan membandingkan metode clustering lain, seperti **DBSCAN atau Hierarchical Clustering**, untuk melihat perbedaan hasil pengelompokan.
3. Mengembangkan sistem agar tidak hanya melakukan *clustering*, tetapi juga **prediksi jumlah penyewaan** menggunakan metode *machine learning* lainnya.

4. Menyempurnakan tampilan antarmuka aplikasi agar lebih interaktif dan mudah digunakan oleh pengguna non-teknis.

DAFTAR PUSTAKA

- <https://www.ibm.com/id-id/think/topics/k-means-clustering>