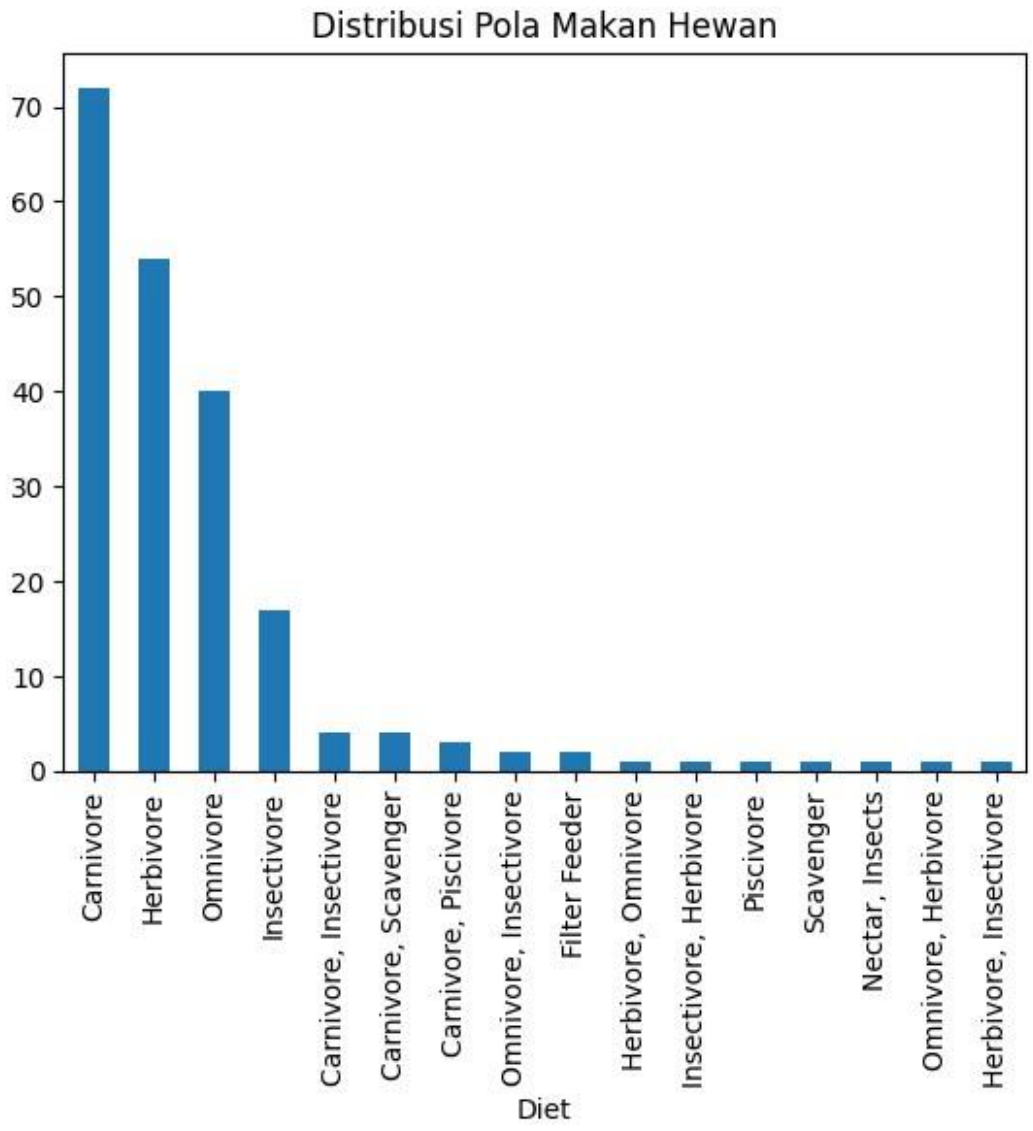
TUGAS UAS - ANALISIS DATA CIRI-CIRI HEWAN MENGGUNAKAN TEKNIK CLUSTERING K-MEANS

```
from google.colab import files
uploaded = files.upload()
<IPython.core.display.HTML object>
Saving Animal Dataset.csv to Animal Dataset.csv
```

ANALISA DATASET import

```
pandas as pd

df = pd.read_csv('Animal Dataset.csv')
df.info() df.describe()
df['Diet'].value_counts().plot(kind='bar', title='Distribusi Pola
Makan Hewan')

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 16 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
0   Animal                   205 non-null    object
1   Height (cm)              205 non-null    object
2   Weight (kg)              205 non-null    object
3   Color                    205 non-null    object
4   Lifespan (years)         205 non-null    object
5   Diet                     205 non-null    object
6   Habitat                  205 non-null    object
7   Predators                205 non-null    object
8   Average Speed (km/h)     205 non-null    object
9   Countries Found          205 non-null    object
10  Conservation Status      205 non-null    object
11  Family                   205 non-null    object
12  Gestation Period (days)  205 non-null    object
13  Top Speed (km/h)         205 non-null    object
14  Social Structure         205 non-null    object 15
    Offspring per Birth      205 non-null    object
dtypes: object(16)
memory usage: 25.8+ KB

<Axes: title={'center': 'Distribusi Pola Makan Hewan'}, xlabel='Diet'>
```



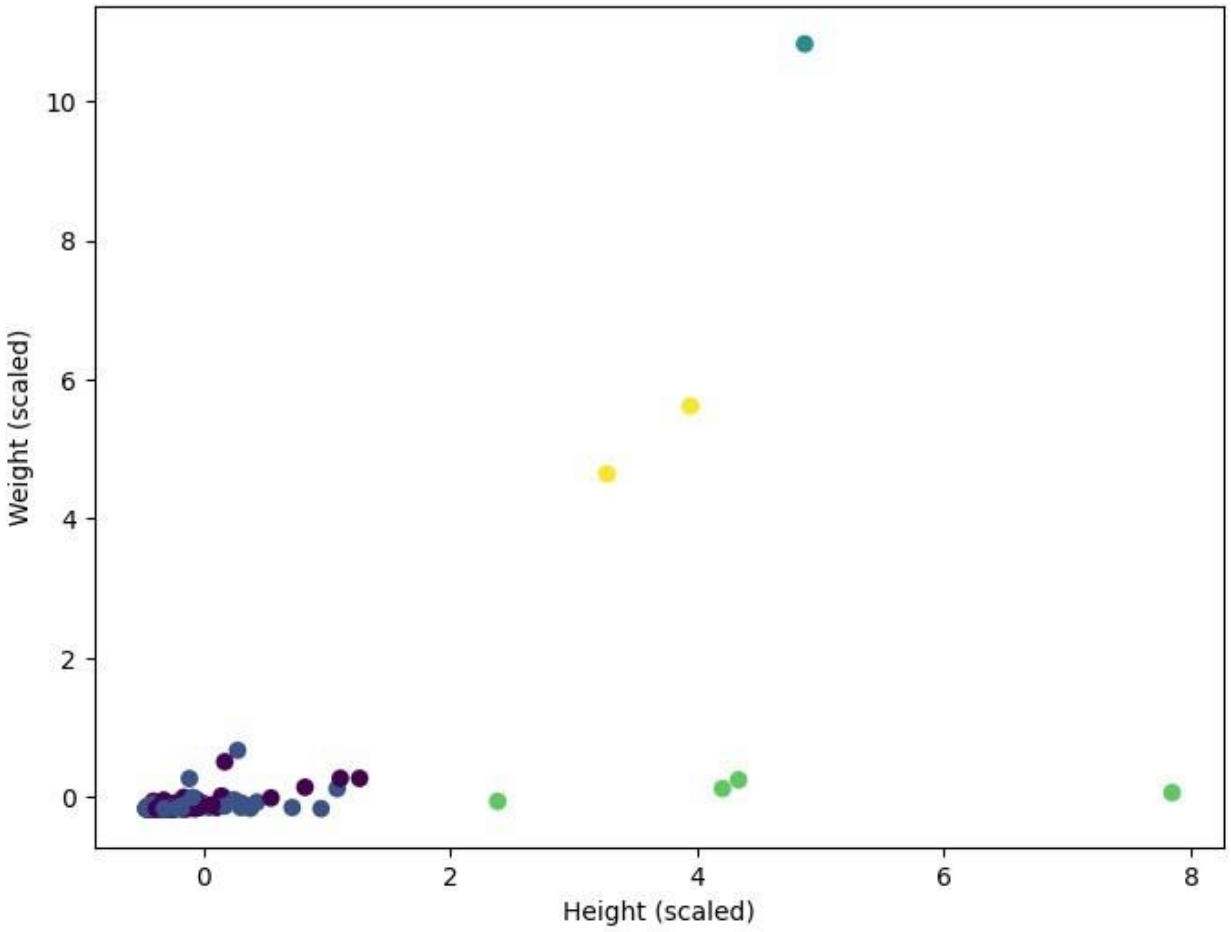
PENGUJIAN MODEL

```
df['Height (cm)'] = df['Height (cm)'].apply(clean_numeric)
df['Weight (kg)'] = df['Weight (kg)'].apply(clean_numeric)
df['Average Speed (km/h)'] = df['Average Speed
(km/h)'].apply(clean_numeric)
df = df.dropna(subset=['Height (cm)', 'Weight (kg)', 'Average Speed
(km/h)'])

X = df[['Height (cm)', 'Weight (kg)', 'Average Speed (km/h)']]
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

kmeans = KMeans(n_clusters=5, random_state=0)
df['Cluster'] = kmeans.fit_predict(X_scaled)
```

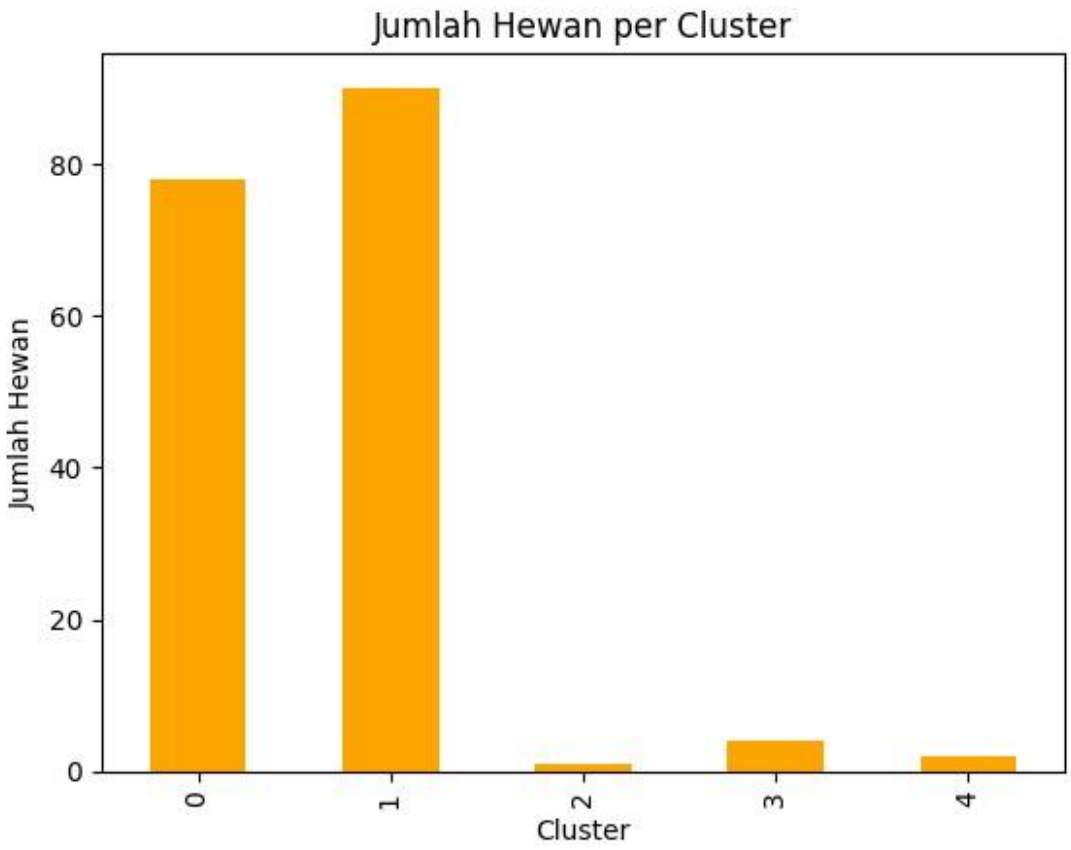
```
df[['Animal', 'Cluster']].head(10)
df['Cluster'].unique() array([0, 1,
3, 4, 2], dtype=int32)
```

VISUALISASI

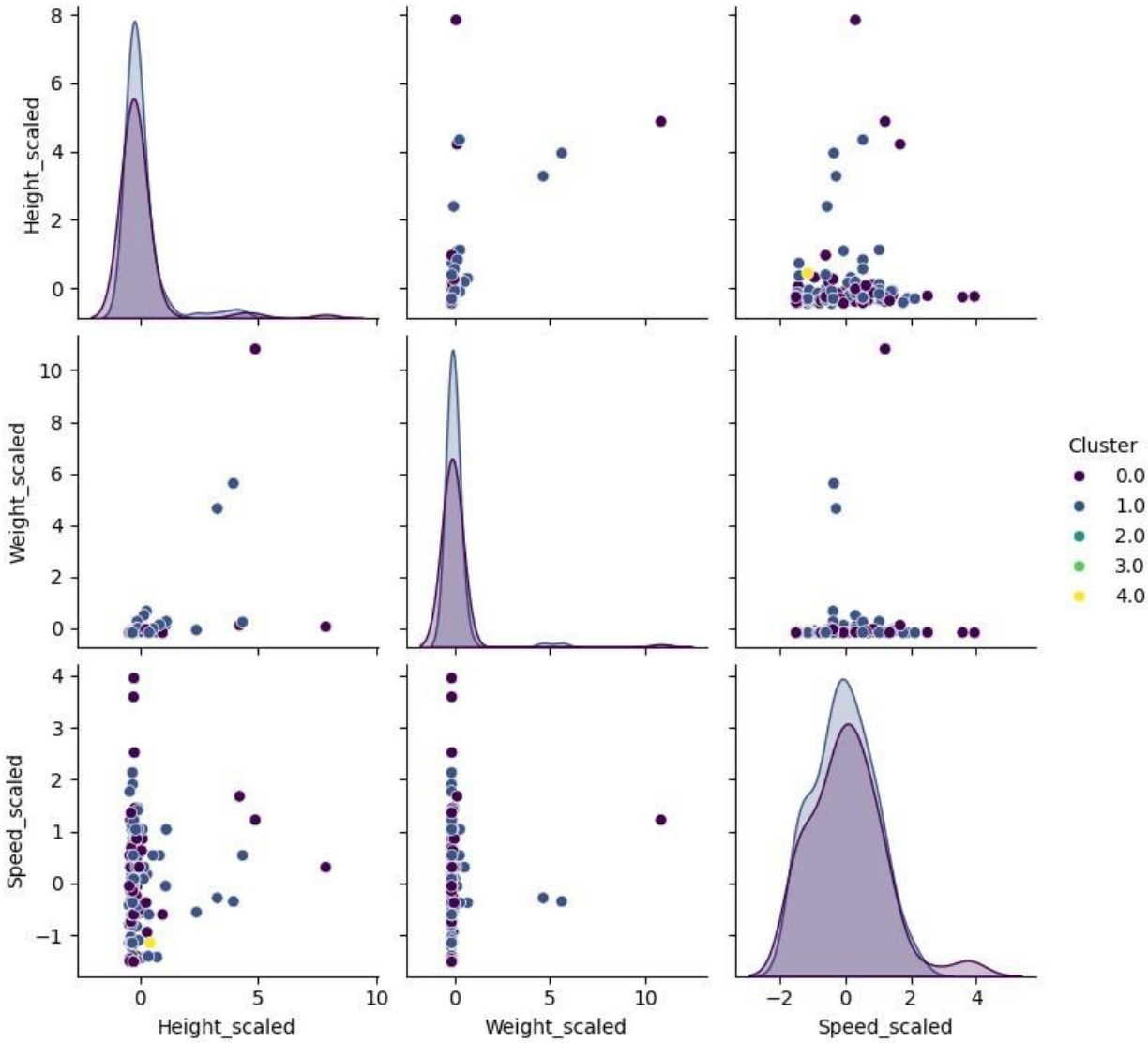
```
import matplotlib.pyplot as plt

plt.figure(figsize=(8,6))
plt.scatter(X_scaled[:, 0], X_scaled[:, 1], c=df['Cluster'],
cmap='viridis')
plt.xlabel('Height (scaled)')
plt.ylabel('Weight (scaled)')
plt.show()
```



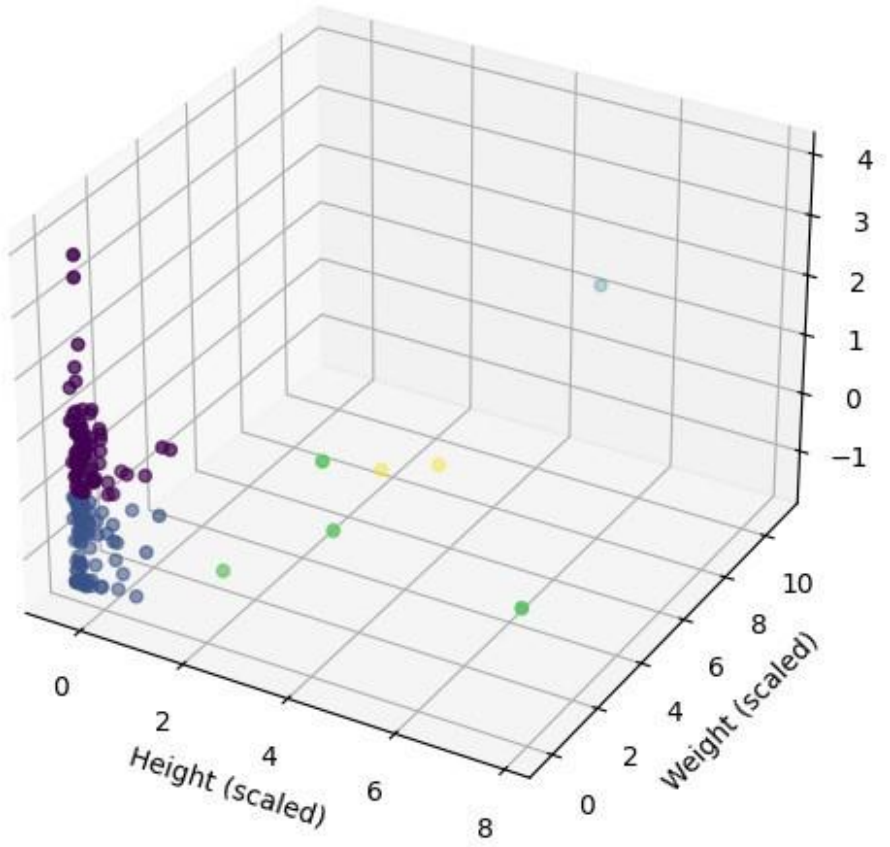

```
df['Cluster'].value_counts().sort_index().plot(kind='bar',
color='orange')
plt.title('Jumlah Hewan per Cluster')
plt.xlabel('Cluster') plt.ylabel('Jumlah Hewan')
plt.show()
```




```
import seaborn as sns

scaled_df = pd.DataFrame(X_scaled, columns=['Height_scaled',
'Weight_scaled', 'Speed_scaled'])
scaled_df['Cluster'] = df['Cluster']

sns.pairplot(scaled_df, hue='Cluster', palette='viridis')
plt.show()
```




```
from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize=(8,6)) ax = fig.add_subplot(111,
projection='3d')
ax.scatter(X_scaled[:, 0], X_scaled[:, 1], X_scaled[:, 2],
c=df['Cluster'], cmap='viridis') ax.set_xlabel('Height
(scaled)') ax.set_ylabel('Weight (scaled)')
ax.set_zlabel('Speed (scaled)') plt.show()
```



Link Github : https://github.com/bryanr34/UAS-Data-Mining

Link colab :
https://colab.research.google.com/drive/1tKK2fVvQWpsalSm7BbFESLksAUsjKc9M?usp=sharing