



KITE Institute of Technology
ARYAN RAJ PANDEY
CSE AI



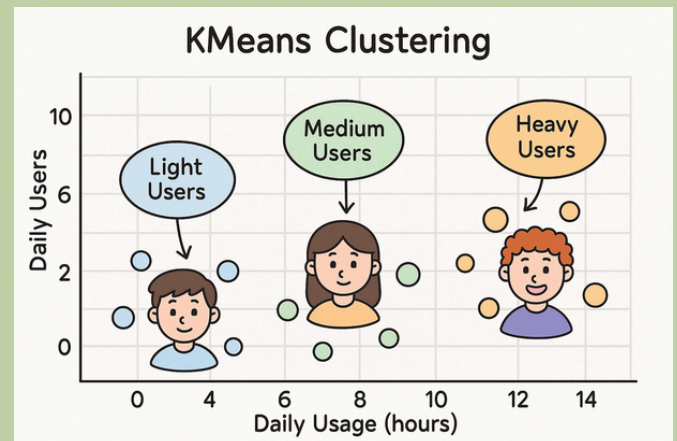
- 2024 / 2025 -

INTERNET USAGE CLUSTERING



2. INTRODUCTION

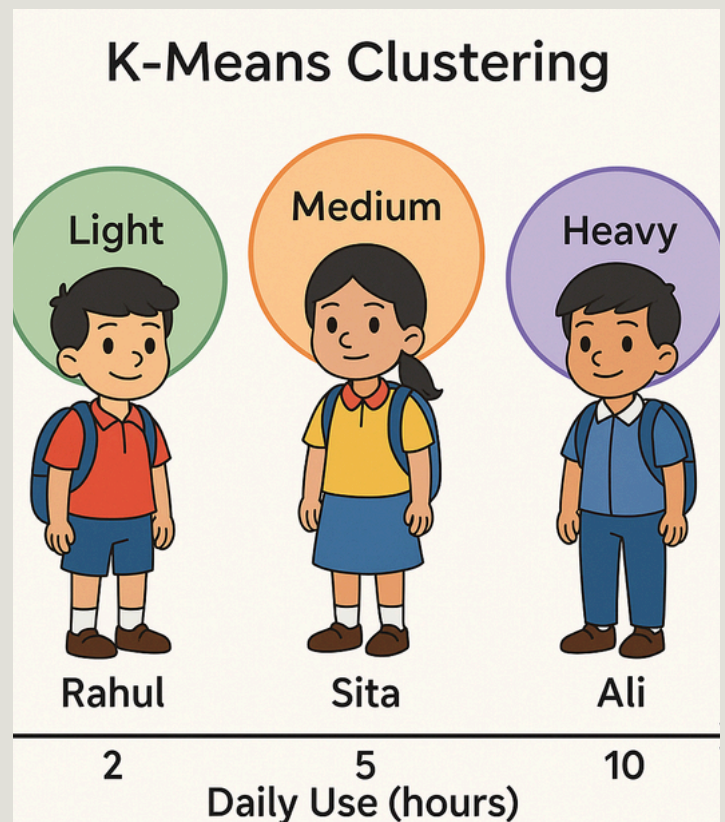
In today's digital age, understanding internet usage patterns can help in categorizing users and personalizing experiences. This project focuses on grouping users based on their online activity data such as daily usage time, site categories visited, and number of sessions per day. By using clustering, we can identify different types of users such as light, medium, and heavy users without any prior labels.



3. Methodology

We used the KMeans Clustering algorithm for this unsupervised learning task. The steps followed in this project were:

- **Data Collection:** Collected data from a CSV file containing user internet usage.
- **Preprocessing:** Standardized the features to bring all values to the same scale.
- **Clustering:** Applied KMeans algorithm to group users into 3 clusters.
- **Visualization:** Used a scatter plot to visualize user clusters based on usage behavior.





4. CODE



```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# DATA LOADING
df = pd.read_csv("internet_usage.csv")

# DATA PROCESSING
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df)

# KMEANS CLUSTERING
kmeans = KMeans(n_clusters=3, random_state=42)
df['cluster'] = kmeans.fit_predict(scaled_data)

# PLOTTING THE CLUSTERING
plt.figure(figsize=(10, 6))
plt.scatter(df['daily_usage_hours'], df['sessions_per_day'],
            c=df['cluster'], cmap='viridis', s=100)
plt.xlabel('Daily Usage Hours')
plt.ylabel('Sessions Per Day')
plt.title('User Clustering Based on Internet Usage')
plt.grid(True)
plt.colorbar(label='Cluster')
plt.show()
```



5. Credits

- Project By: ARYAN RAJ PANDEY
 - Guided By: BIKKI KUMAR
 - Tools Used: Python, Pandas, Scikit-learn, Matplotlib
 - Data Source: internet_usage.csv
-