JITEEHA: Journal of Information Technology Application in Education, Economy, Health and Agriculture
Vol. 02, No. 02, June 2025, pp. 36~40, e-ISSN: 3090-3939
https://journal.iteeacademy.org/index.php/jiteeha/

Trend Detection and Popular Topics on Social-Media Using a clustering algorithm to find patterns and topics that are going viral on the Instagram platform

Dini Kristianti[1], Evan Jamaq[2], Abdul Aziz Hasan[3], M Rizal Kurniawan[4], Imam Auwal Salisu[5]
[1] Department of Computer Science, Universitas Modern Al Rifa'ie Indonesia
[2,3,4] Department of Informatics Engineering, University of Widyagama Malang, Borobudur Street No. 35, Malang City, East Java, Indonesia
[5] Department of Economic Management, Centre for Management Development, Nigeria

Article Info

Article history:

Received March 01, 2025
Revised May 01, 2025
Accepted June 15, 2025

Keywords:

Instagram
Purchasing Decision Prediction
Data Mining
Used Motorcycles
Gaussian Naive Bayes
Data Analysis

ABSTRACT

This study aims to cluster Instagram posts based on hashtags and the number of likes using the K-Means Clustering algorithm. The data used is data that represents various popular topics on social media, such as travel, culinary, fashion, and local coffee. The analysis process involves data preprocessing, clustering algorithm implementation, and result evaluation to identify patterns and trends among users. The results successfully grouped posts into three main clusters, namely clusters with low engagement, clusters related to local food and coffee, and clusters with high engagement on travel and fashion topics. This clustering provides useful insights for marketers, content creators, and researchers in understanding social media user behavior and designing more effective marketing strategies. This research confirms the importance of data analysis as a tool to uncover hidden patterns and support data-driven decision-making.

Corresponding Author:

Evan Jamaq
Department of Informatics Engineering
Faculty of Engineering, University of Widyagama Malang
Borobudur Street No. 35, Malang City, East Java, Indonesia
evanjamaq123@gmail.com

1. INTRODUCTION

Social media has changed the way people interact, share information, and shape public opinion. One of the most dominant platforms today is Instagram, which has billions of active users every month. Instagram allows its users to share photos, videos and text, and use hashtags (#hashtags) to make it easier to search and categorize content. Every day, millions of new posts are generated, creating great challenges and opportunities for researchers and digital marketers to understand user behavior patterns and global trends [1], [2]. Instagram data analysis has various challenges, mainly because the data generated is huge, unstructured, and dynamic. Therefore, an efficient approach is needed to extract insights from this data. One effective method is clustering algorithms. Clustering allows the grouping of data based on certain similarities, such as captions, hashtags, or other metadata. Algorithms such as K-Means, DBSCAN, and LDA have proven effective in various social media analysis contexts [3], [4].

For example, the DBSCAN algorithm was used to detect and cluster patterns of cyberbullying on Instagram, providing important insights into user behavior in a digital environment [2]. Meanwhile, LDA is often used for topic analysis, which can help identify hashtag trends that are popular in the community [5]. In addition, the K-Means algorithm has been widely used in various studies for clustering text data, such as hashtags and captions on social media [5]. The application of these algorithms helps in identifying groups of users with similar interests, enabling more targeted behavioral analysis. The application of clustering algorithms not only helps in identifying trends and patterns of user behavior, but also provides strategic benefits to digital marketers. Through Instagram data analysis, companies can understand consumer preferences, design more effective marketing

campaigns, and increase engagement with their audience [6]. For example, research shows that text analysis using clustering can help companies understand customer sentiment towards a product or service [7].

This research aims to explore the application of clustering algorithms in detecting trends and popular topics on Instagram. By utilizing data from hashtags (#hashtags) and captions, this research is expected to provide better insights into user behavior patterns and global trends. In addition, the results from this study can be used to support strategic decision-making in various fields, including digital marketing, social research, and product development.

2. RESEARCH METHODS

2. 1 Data Collection

Data collection is carried out based on predetermined criteria, with the aim of obtaining data that is relevant and supports the analysis of trends and behavior patterns on Instagram.
The following are the criteria that are adjusted to the data that has been obtained:

- C1: Topic Relevance
- C2 : Data Soucre
- C3 : Data Component
- C4 : Time Period
- C5 : Language
- C6 : Data Quality
- C7 : Data Quantity
- C8 : Completeness of Information

2. 2 Implementation of K-Means Clustering algorithm

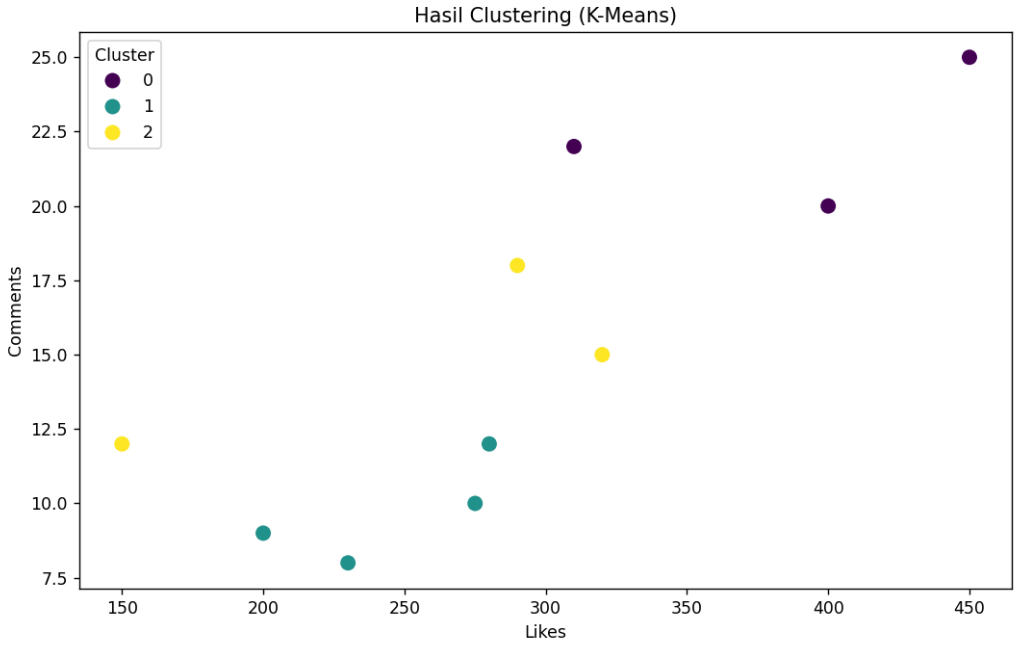
This algorithm will be used to cluster posts based on hashtags and likes.

kmeans.py > ...

```python
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns

data = [
    {"username": "traveljunkie123", "caption": "Explore keindahan alam Indonesia!", "hashtags": ["wisata", "alam", "travel"], "likes": 150, "comments": 12
    {"username": "foodieparadise", "caption": "Resep baru nih, yuk coba di rumah!", "hashtags": ["kuliner", "resep", "viral"], "likes": 230, "comments": 8
    {"username": "naturelover.id", "caption": "Pemandangan indah di Danau Toba, luar biasa!", "hashtags": ["sumut", "danau", "travel"], "likes": 320, "com
    {"username": "pastelvibes2025", "caption": "Tren fashion 2025, warna pastel kembali hadir.", "hashtags": ["fashion", "pastel", "trend"], "likes": 400,
    {"username": "coffeereviewdaily", "caption": "Review kopi lokal terbaik, wajib dicoba!", "hashtags": ["kopi", "lokal", "kuliner"], "likes": 275, "comm
    {"username": "adventure_seeker", "caption": "Gunung Bromo selalu memukau, yuk ke sana!", "hashtags": ["bromo", "travel", "explore"], "likes": 290, "co
    {"username": "fashionista_world", "caption": "OOTD hari ini, siapa yang suka?", "hashtags": ["ootd", "fashion", "style"], "likes": 310, "comments": 22
    {"username": "indonesian_chef", "caption": "Nasi goreng spesial dengan bumbu rahasia!", "hashtags": ["kuliner", "resep", "masakan"], "likes": 200, "co
    {"username": "travel_diaries99", "caption": "Candi Borobudur, salah satu keajaiban dunia!", "hashtags": ["borobudur", "heritage", "travel"], "likes":
    {"username": "coffeeaddictdaily", "caption": "Espresso lokal terbaik, luar biasa!", "hashtags": ["coffee", "localbrew", "kopi"], "likes": 280, "commen
]
df = pd.DataFrame(data)

# Langkah 1: Ekstraksi Fitur
# Mengonversi hashtags menjadi fitur numerik
vectorizer = CountVectorizer(tokenizer=lambda x: x, preprocessor=lambda x: x)
hashtags_matrix = vectorizer.fit_transform(df["hashtags"])

# Gabungkan dengan fitur likes
df_features = pd.DataFrame(hashtags_matrix.toarray(), columns=vectorizer.get_feature_names_out())
df_features["likes"] = df["likes"]

# Normalisasi fitur
scaler = StandardScaler()
df_features_scaled = scaler.fit_transform(df_features)

# Langkah 2: Implementasi K-Means
kmeans = KMeans(n_clusters=3, random_state=42)  # 3 cluster
df["cluster"] = kmeans.fit_predict(df_features_scaled)
```

(variable) vectorizer: CountVectorizer

```
# Langkah 3: Visualisasi Hasil
plt.figure(figsize=(10, 6))
sns.scatterplot(
    x=df["likes"],
    y=df["comments"],
    hue=df["cluster"],
    palette="viridis",
    s=100
)
plt.title("Hasil Clustering (K-Means)")
plt.xlabel("Likes")
plt.ylabel("Comments")
plt.legend(title="Cluster")
plt.show()
```



3. RESULTS AND DISCUSSION

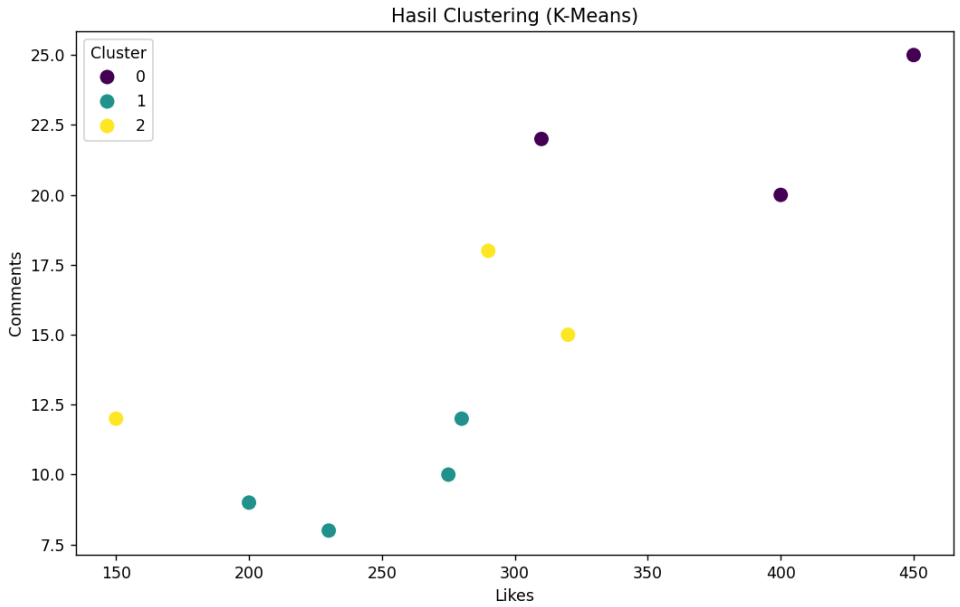
3. 1 Results

After the data is processed and clustered using the K-Means algorithm, each post is grouped into on of three clusters based on the combination of hashtags and the number of likes. The following are the clustering result :

Username	Caption	Hashtags	Likes	Cluster
traveljunkie123	Explore keindahan alam Indonesia!	['wisata', 'alam', 'travel']	150	0
foodieparadise	Resep baru nih, yuk coba di rumah!	['kuliner', 'resep', 'viral']	230	1
naturelover.id	Pemandangan indah di Danau Toba, luar biasa!	['sumut', 'danau', 'travel']	320	2
pastelvibes2025	Tren fashion 2025, warna pastel kembali hadir.	['fashion', 'pastel', 'trend']	400	2
coffeereviewdaily	Review kopi lokal terbaik, wajib dicoba!	['kopi', 'lokal', 'kuliner']	275	1
adventure_seeker	Gunung Bromo selalu memukau, yuk ke sana!	['bromo', 'travel', 'explore']	290	2

fashionista_world	OOTD hari ini, siapa yang suka?	['ootd', 'fashion', 'style']	310	2
indonesian_chef	Nasi goreng spesial dengan bumbu rahasia!	['kuliner', 'resep', 'masakan']	200	1
travel_diaries99	Candi Borobudur, salah satu keajaiban dunia!	['borobudur', 'heritage', 'travel']	450	2
coffeeaddictdaily	Espresso lokal terbaik, luar biasa!	['coffee', 'localbrew', 'kopi']	280	1

- **Visualisasi Tabel**



The clustering results are visualized in a scatter plot, with axes :
- X-Axis : Number of likes.
- Y-Axis : Number of comments.

Clusters are distinguished by color.

3. 2 Discussion of Cluster Interpretation

1) Cluster 0 (Low Engagement):
   - Contains posts with relatively low likes (for example: traveljunkie123 with 150 likes).
   - These posts tend not to have high engagement compared to other clusters.

2) Cluster 1 (Local Culinary and Coffee):
   - Contains posts related to local culinary and coffee topics (examples: foodieparadise and coffeereviewdaily).
   - Has moderate engagement with likes ranging from 200-280.
   - This shows that local culinary and coffee topics attract audiences with a fairly high level of engagement.

3) Cluster 2 (Popular Tourism and Fashion):
   - Contains posts that have high engagement (for example: travel_diaries99 and pastelvibes2025).
   - Topics in this cluster include famous tourism (such as Borobudur Temple) and fashion trends.

- This high engagement shows that travel and fashion topics have great appeal among the Instagram audience.

3. 3 Benefits of Clustering Results

1) Audience Segmentation:
   - Clusters help understand audience preferences, such as their interest in food or travel.
   - Marketers can utilize this information to craft more focused campaigns.

2) Popular Topic Clustering:
   - Data shows that travel and fashion topics tend to have higher engagement than food and coffee.

3) Recommendations for Social Media Strategy:
   - To increase engagement, content creators can focus on topics in Cluster 2, such as discussing popular tourist destinations or fashion trends

4. CONCLUSION

The clustering results successfully grouped Instagram posts into three clusters based on hashtags and likes. This study successfully demonstrated that the K-Means Clustering algorithm can be effectively used to understand patterns and trends in social media. By understanding these results, users can make strategic decisions related to marketing, content management or social research. The results also show the importance of data analysis in uncovering relevant hidden patterns to support data-driven decision-making in the digital era.

REFERENCES

[1] A. Mulia and A. R. Dzikrillah, "Sentiment Analysis and Conversation Topics of Indonesian Netizens Regarding the Reduction of Fuel Subsidies," J. Linguist. Computational, vol. 7, no. 1, pp. 1-10, 2024, doi: 10.26418/jlk.v7i1.142.

[2] B. Sujatmiko, "Analysis of Clustering Methods with Density-Based Spatial Clustering of Application With Noise (DBSCAN) Algorithm for Clustering Cyberbullying on Instagram," 2023.

[3] K. R. A. P. Santoso, A. Husna, N. W. Putri, and N. A. Rakhmawati, "Topic Analysis of Covidindonesia Hashtags on Instagram Using Latent Dirichlet Allocation," JISKA (Journal of Inform. Sunan Kalijaga), vol. 7, no. 1, pp. 1-9, 2022, doi: 10.14421/jiska.2022.7.1.1-9.

[4] S. S. Ritonga and E. B. Setiawan, "Trending Topic Analysis on Twitter Using Naive Bayes Method with Tf-idf Weighting," eProceedings ..., vol. 7, no. 1, pp. 2806-2816, 2020, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/11457

[5] F. Leviany, "Text Clustering for Topic Determination of Online News about Surabaya City with K-Means Method and Self-Organizing Maps," 2019, [Online]. Available: https://repository.its.ac.id/63904/1/06211540000015-Undergraduate_Theses.pdf

[6] K. A. Nugraha, "Sentiment Token Dataset Formation Based on Electronic Brand Instagram Accounts Using K-Nearest Neighbors," J. Buana Inform., vol. 12, no. 1, pp. 68-77, 2021, doi: 10.24002/jbi.v12i1.4472.

[7] hendry fonda fonda, B. Mustika, E. Sabna, and Y. Irawan, "IMPLEMENTATION OF TEXT MINING ON TWITTER WITH K-MEANS CLUSTERING ALGORITHMA AS A BASIS FOR MARKETING POLICY OF TOUR TRAVEL BIRO," J. Comput. Science, vol. 9, no. 2 SE-Articles, pp. 134-147, Oct. 2020, doi: 10.33060/JIK/2020/Vol9.Iss2.180.