Received April 11th, 2025; accepted June 18th, 2025. Date of publication June 30th, 2025
Digital Object Identifier: https://doi.org/10.25047/jtit.v12i1.443

Clustering-based Machine Learning Approach For Predicting Tourism Trends From Social Media Behavior

CANDRA AGUSTINA, EKA RAHMAWATI

Universitas Bina Sarana Informatika, Jakarta Indonesia

CORESPONDING AUTHOR: CANDRA AGUSTINA (email:candra.caa@bsi.ac.id)

ABSTRACT Digital technology has significantly transformed tourist behavior, particularly in searching for, selecting, and sharing travel experiences. Social media has become a primary source of information, influencing travel decisions through real-time recommendations and user-generated content. However, the large volume of data generated by social media presents challenges in understanding and predicting tourist behavior. This study aims to analyze tourist behavior patterns using a clustering-based machine learning approach, specifically K-Means Clustering. The research examines engagement levels on platforms such as Instagram, TikTok, and TripAdvisor to categorize tourists into three key segments: Digital-Savvy Travelers, Passive Travelers, and Conservative Travelers. The results indicate that machine learning effectively analyzes large-scale tourism data, providing valuable insights for destination marketing, personalized recommendations, and service optimization. The findings highlight the potential of machine learning to identify emerging trends, improve customer segmentation, and enhance targeted promotional strategies. Understanding these patterns enables tourism businesses to create data-driven strategies aligned with modern travel behaviors. In a broader perspective, artificial intelligence can revolutionize tourism marketing, increase customer engagement, and improve the overall travel experience.

KEYWORDS: social media; tourist behavior; machine learning; k-means clustering; travel decision

I.INTRODUCTION

Digital technology has changed how tourists search, select, and share their experiences regarding tourist destinations. Social media is one of the immediate bases of information for tourists in planning their trips, from looking for destination ideas and ensuring the right choice to getting information on accommodations and activities that can be done at the destination. According to research Tanković et al., (2022), tourists increasingly rely on social media to obtain faster, more manageable, and more up-to-date information than conventional sources [1]. Apart from that, communication in social networks also plays an essential role in increasing tourists' insight into a destination through experiences shared by other users.

The increasing social media usage in the tourism sector poses new challenges to understanding tourist behavior. The data on social media is very diverse and reflects tourist interaction patterns in several forms, such as searching for destination information, activity preferences, and communication patterns used in sharing tourism experiences [2]. With substantial data volumes, conventional methods of Analyzing tourism trends have become less effective. Therefore, an artificial intelligence-based approach, especially Clustering-based Machine Learning, is needed to group tourist behavior patterns more accurately [3].

Clustering techniques in Machine Learning allow tourists to be grouped based on information search characteristics, type of communication used, and behavioral patterns in sharing tourism content on social media. This approach predicts tourism trends based on interaction patterns on several digital platforms, such as Instagram, TikTok, Twitter, Facebook, and TripAdvisor. The use of clustering in Analyzing tourist behavior on social media can help tourism industry players develop more effective marketing strategies, identify destinations that are on the rise, and improve service quality based on user preferences.

This research aims to develop a tourism trend prediction model using a Clustering-based Machine Learning approach by Analyzing tourist behavior using social media. It is hoped that the results will provide deeper insight into tourist travel patterns and support decision-making for stakeholders in the tourism industry.

1.1 The Influence of Social Media on Tourist Behavior

Social media has become a key platform for travelers to gather information, share experiences, and influence travel decisions. According to Lama, the rapid growth of social media in the tourism industry has changed how tourists interact with destinations, service providers, and fellow travelers. Platforms like Instagram, Facebook, TikTok, and Twitter enable real-time interactions, allowing potential travelers to explore destinations through user-generated content, reviews, and multimedia experiences [4]. The ease of access and instant nature of social media give travelers the ability to make more informed decisions by evaluating experiences shared by other users, which significantly shapes their travel plans and expectations [5, 6, 7]. Additionally, integrating social media in digital marketing has enabled tourism businesses to effectively target and interact with their audiences, creating a more dynamic and interconnected travel ecosystem.

Besides influencing individual decisions, social media also plays a vital role in destination marketing strategies. Integrating social media into digital marketing allows tourism businesses to reach a broader audience in a more personalized way. Through data-driven marketing strategies, companies can target travelers based on their preferences, search history, and interactions on digital platforms [8]. This makes travel more dynamic, where tourists act as consumers and content creators who help shape the image of a destination in the digital realm [9]. Thus, social media is not just a communication tool but also an instrument that redefines how the tourism industry develops in the digital era [10].

1.2 Clustering in Social Media Data Analysis

Clustering is an unsupervised learning technique that aims to group data into several groups based on pattern similarities [11]. This method works without data labels, so the algorithm must automatically find natural structures in the data. Clustering is frequently used in several domains, including image analysis, pattern recognition, customer analysis, and travel recommendation systems, due to its ability to identify hidden patterns that traditional methods cannot discover.

According to Lad & Metkewar (2020), Clustering techniques have been applied in several technological and industrial fields. In speech recognition, clustering groups sounds based on acoustic features to improve speech recognition accuracy. In face detection, this method helps identify and classify faces in security systems and biometric recognition. Clustering is also applied in spam detection, where algorithms can automatically group suspicious emails based on word patterns and senders. In the financial sector, credit card fraud detection utilizes clustering to recognize unusual transactions as an indication of fraud. In addition, in medical diagnosis, this technique helps group patients based on symptoms and laboratory results to improve early disease detection and personalized treatment [12, 13].

One algorithm that is often used in clustering is K-Means Clustering. This algorithm is an unsupervised learning technique that groups data into several clusters based on pattern similarity and uses a distance-based approach to determine the proximity between data in each cluster. According to Wu (2021) in the journal K-Means Clustering Algorithm and Python Implementation, K-Means is the algorithm most commonly used in machine learning because of its simplicity and efficiency in grouping data into several categories [14].

The K-Means algorithm determines the number of clusters (K) first, then iteratively assigns each data point to the closest cluster based on Euclidean distance. After all points are categorized, the algorithm recalculates the cluster centers (centroids). It repeats the process until there is no significant change in the division of clusters or until a convergence condition is reached[15], [16]. One of the main advantages of this method is its speed because the K-Means algorithm can work with large amounts of data and high dimensions with relatively high efficiency [17, 18, 19, 20].

Social media has become a vast and complex data source, with millions of users interacting daily across platforms such as Twitter, Facebook, Instagram, and LinkedIn. The data generated from these interactions can be analyzed using clustering techniques to group specific users, content, or trends [21, 22]. Applying this technique allows a deeper understanding of user behavior, audience segmentation, and data-based optimization of marketing strategies [23].

According to Zhurakovskyi et al. (2023), social networking platforms are a rich source of information for Analyzing user interaction patterns. Data clustering methods, such as K-Means and Mini Batch K-Means, are used to group user behavior and preferences on a large scale. This approach makes marketing strategies more effective and personalized [24]. This technique also plays a role in detecting market trends and increasing content distribution on social media. Thus, cluster-based analysis strengthens user engagement and provides broader knowledge for stakeholders.

Although various advanced clustering methods such as DBSCAN, Hierarchical Clustering, and Spectral Clustering have been explored in recent studies, this research employs K-Means Clustering due to its simplicity, interpretability, and effectiveness in handling structured survey data. This study specifically addresses the gap in segmenting tourist behavior patterns using social media interaction metrics in the Indonesian context, which remains underexplored in recent literature.

1.3 Clustering Evaluation

Clustering was used for the evaluation. Clustering is a technique in unsupervised learning that aims to group data into several clusters based on the similarity of certain patterns. An evaluation method is needed to determine the quality of the clusters formed.

One commonly used method is the Elbow Method, which helps determine the optimal number of clusters in K-Means Clustering. This method works by calculating the Within-Cluster Sum of Squares (WCSS) for several values of K (number of clusters). WCSS measures the sum of the squared distances between each data point and the specified cluster centroid [25].

II.METHOD

2.1 Research Framework

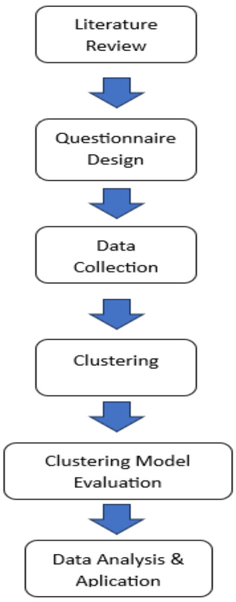

FIGURE 1. Research Framework

Figure 1 shows that this research began with a literature study to understand previous research on K-Means Clustering and social media-based tourist behavior analysis. This study aims to identify theories, central concepts, and research gaps as a basis for research. The next stage is designing the questionnaire, where the research instrument is prepared based on the variables that have been identified. This questionnaire is used to collect primary data from relevant respondents. Trials were carried out to ensure its validity and reliability before full-scale deployment.

After the design was completed, data was collected through questionnaires distributed to respondents. The collected data was then processed and prepared for further analysis. Before the clustering process is carried out, the data collected through questionnaires undergoes a preprocessing stage, including data cleaning, handling missing data, and normalizing numerical variables to ensure data consistency [26, 27, 28]. After that, the K-Means Clustering algorithm was applied to group respondents into several clusters based on their answer patterns.

The elbow method is used to determine the optimal number of clusters (K). This method calculates the Within-Cluster Sum of Squares (WCSS) for several values of K and plots the results. The elbow point, where the decline in WCSS begins to slow down, is chosen as the most suitable number of clusters. This clustering process uses Python in Visual Studio Code (VS Code). Data was processed using the pandas library for data management, scikit-learn for implementing K-Means Clustering, and matplotlib for Elbow Method visualization [29, 30]. This approach ensures that social media-based tourist segmentation analysis is conducted efficiently and reproducibly.

The final stage is data analysis and interpretation, where the clustering results are analyzed to understand tourist behavior patterns. These results provide recommendations for the tourism industry to develop data-based marketing strategies and a more effective tourist destination recommendation system.

2.2 Dataset

The data collection process involved distributing an online questionnaire to 350 respondents who actively use social media for travel-related purposes. The questionnaire was designed based on prior literature and comprised three key sections:

1. Use of Social Media for Travel Planning
   Respondents were asked to indicate how frequently they use social media to get travel ideas, search for destination information, confirm their choice of destination, look for accommodation, and explore activities at the destination.
2. Social Media Communication
   This section assessed how respondents perceive speed, ease of use, and knowledge enhancement through social media interactions.
3. Social Media Information Quality
   Respondents evaluated aspects such as efficiency, availability of unique information, ease compared to other media, improvement in information quality, accessibility, timeliness, and up-to-date nature of information [1].

All items in the questionnaire were measured using a Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) to quantify respondents' perceptions.

III.RESULT AND DISCUSSION

3.1 Data Preprocessing

The first step in implementing K-Means Clustering is data preprocessing. which includes:

1. Read the dataset from the social.csv file using the pandas library.
2. Handle missing values with the drop() method to ensure the data used is complete.
3. Normalize the data using StandardScaler from the scikit-learn library so that each variable has a comparable scale, making it easier for the clustering algorithm to detect patterns.

The program listing is shown in Table 1.

TABLE 1. Listing Program of Preprocessing

```
import pandas as pd; import numpy as np
import matplotlib.pyplot as plt; import pickle
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
file_path = "social.csv"
df = pd.read_csv(file_path, sep=';', engine='python')
df = df.apply(pd.to_numeric, errors='coerce')
df = df.dropna()
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df)
wcss = []
K_range = range(1, 11)  # Mencoba K dari 1 hingga 10
for k in K_range:
    kmeans = KMeans(n_clusters=k, init='k-means++', max_iter=300,
n_init=10, random_state=42)
    kmeans.fit(df_scaled)
    wcss.append(kmeans.inertia_)
plt.figure(figsize=(8, 5))
plt.plot(K_range, wcss, marker='o', linestyle='-')
plt.xlabel("Number of Clusters (K)")
plt.ylabel("WCSS (Within-Cluster Sum of Squares)")
plt.title("Elbow Method for Optimal K")
plt.xticks(K_range)
plt.grid(True)
plt.show()
kmeans = KMeans(n_clusters=3, init='k-means++', max_iter=300,
n_init=10, random_state=42)
df['Cluster'] = kmeans.fit_predict(df_scaled)
df.to_csv("sosial_clustered.csv", index=False)
with open("kmeans_model.pkl", "wb") as model_file:
    pickle.dump(kmeans, model_file)
with open("scaler.pkl", "wb") as scaler_file:
    pickle.dump(scaler, scaler_file)
```

Table 1 contains the program script to create a cluster, which is run in VS Code and saves the model in the kmeans_model.pkl file for further processing.

3.2 Determine the Number of Clusters using the Elbow Method

The Elbow Method determines the optimal number of clusters (K). This method works by calculating the Within-Cluster Sum of Squares (WCSS) for several values of K and plotting the results. From the visualization results displayed in the Elbow Method Figure, the WCSS value decreases significantly up to the elbow point, which is at K = 3, indicating that three clusters are the optimal number for this research.

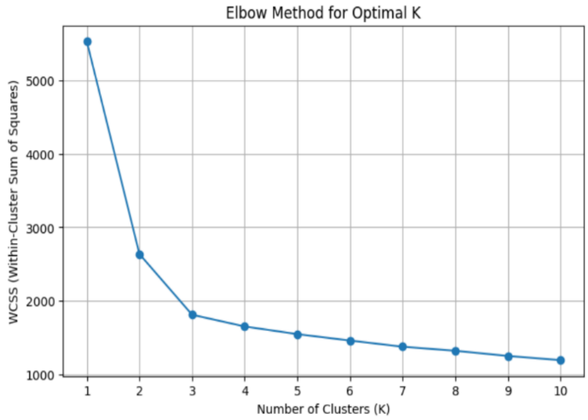

FIGURE 2. Elbow graph

Based on Figure 2 of the Elbow Method results, it can be concluded that the optimal number of clusters in this analysis is K = 3. The graph shows that the Within-Cluster Sum of Squares (WCSS) value experiences a sharp decrease from K = 1 to K = 2, and continues to decrease until K = 3. However, after K = 3, the decrease in WCSS begins to slow down, which indicates that adding more clusters will not make a significant difference in data segmentation. This point, known as the "elbow point", is the main indicator in determining the most optimal number of clusters.

By using K = 3, data can be grouped efficiently without significant loss of information. Determining these clusters allows a better understanding of tourist behavior patterns based on social media, with each cluster representing certain characteristics in the use, communication and search for tourist information. Therefore, the K-Means Clustering process with K = 3 will be applied in the next analysis to group respondents into three main segments that have similar patterns.

3.3 Implementation of K-Means Clustering

After the optimal number of clusters is determined, K-Means Clustering is applied with K = 3. This algorithm works by grouping respondents based on similarities in social media usage patterns. The clustering results show that respondents can be grouped into three main segments, which will be analyzed further to understand the characteristics of each cluster. Clustering results are stored in social cluster.csv and social_klaster.pkl.

Cluster 0 - Digital-Savvy Travelers

Characteristics:

1. Active Engagement: These travelers frequently use social media for quick communication and real-time information [31].
2. Community Interaction: They actively engage with travel communities or online travel groups, sharing experiences and seeking advice.
3. Information Sharing: They are likely to share their travel experiences and content on social media platforms, contributing to user-generated content [32].
4. Influence on Decision-Making: Their travel decisions are heavily influenced by social media

content, including user reviews and recommendations [33].

Cluster 1 - Passive Travelers

Characteristics:

1. Information Consumers: These travelers rely on social media primarily for reading information without much interaction.
2. Preference for Blogs and Forums: They prefer gathering information from blogs or discussion forums rather than actively participating in discussions.
3. Limited Sharing: They are less likely to share their own travel experiences or engage in content creation.
4. Decision-Making: Their travel decisions are influenced by the information they consume passively, such as reviews and recommendations, but they do not actively seek to influence others.

Cluster 2 - Conservative Travelers

Characteristics:

1. Minimal Social Media Use: These travelers use social media sparingly and are less influenced by digital content [34].
2. Traditional Information Sources: They rely more on traditional sources of information, such as travel agencies or printed guides.
3. Cautious Decision-Making: Their travel decisions are less likely to be swayed by social media trends or user-generated content.
4. Cultural Influence: Their behavior may be influenced by cultural values, leading to a preference for familiar and trusted sources of information.

TABLE 2. Comparison Results

Cluster	Engagement Level	Information Sources	Content Sharing	Influence on Decision-Making
Digital-Savvy Travelers	High	Social media, travel communities	High	Strong
Passive Travelers	Low	Blogs, forums	Low	Moderate
Conservative Travelers	Minimal	Traditional sources	Minimal	Low

Table 2 compares three tourist clusters based on their engagement level, information sources, content-sharing behavior, and influence on decision-making. Digital-Savvy Travelers are highly engaged, rely on social media and travel communities, actively share content, and have a strong influence on travel decisions. Passive Travelers have low engagement, gather information from blogs and forums, rarely share content, and are moderately influenced by online sources. Conservative Travelers show minimal engagement, depend on traditional sources, rarely share content, and are the least influenced by digital platforms. This classification helps tourism businesses tailor marketing strategies to different traveler behaviors.

The findings of this study are consistent with previous research by Tanković et al. (2022), who also found that social media engagement significantly impacts travel decisions. However, unlike studies that focused only on text analysis from a single platform such as Twitter, this research incorporates multiple platforms (Instagram, TikTok, TripAdvisor) and behavioral metrics, providing a broader perspective on tourist segmentation.

3.3.1 Key Insights on Social Media Influence on Travel Behaviors

Social media plays a crucial role in shaping travel behaviors at various stages of the travel journey. The following sections explore how social media influences travelers' decision-making processes before, during, and after the trip.

Pre-Trip Planning:

1. Information Search and Decision Making
   Travelers often use social media to gather information and make travel decisions. User-generated content (UGC) on platforms like Facebook, Instagram, and YouTube significantly influences destination choices and travel plans [5], [35, 36]. Social media provides a trustworthy source of information, helping travelers to plan better and make informed decisions.
2. Seeking Feedback
   Before traveling, individuals frequently seek feedback from their social networks, which can shape their travel plans and expectations.

During the Trip:

1. Experience Sharing: Travelers share their experiences in real-time through posts, photos, and videos, which can influence their social circles and enhance their own travel experiences [37]. This sharing behavior is driven by factors such as perceived enjoyment, authenticity, and the desire to document memories.
2. Emotional and Perceptual Impact: Sharing positive experiences can increase travelers' positive affect and overall trip satisfaction, while sharing negative experiences can help mitigate negative perceptions [38].

Post-Trip Sharing:

1. Influence on Others: Post-trip sharing of travel experiences continues to influence the travel decisions of others. Positive reviews and engaging content can enhance the destination's image and attract future travelers [39].
2. Building Trust and Community: Frequent travelers consider the needs and reactions of their followers, aiming to foster a sense of community and improve personal ties through their sharing behavior [40].

3.3.2 Tailoring Marketing Strategies

To effectively tailor marketing strategies and communication efforts, it is crucial to recognize the distinct patterns of social media use across different traveler segments:

1. Targeted Content
   Develop content that resonates with the specific needs and preferences of different traveler groups. For example, visually appealing and engaging content can attract younger travelers, while detailed and informative posts may appeal to more cautious or information-seeking travelers [41].
2. Leveraging Influencers

Utilize social media influencers (SMIs) to promote destinations. SMIs can build trust and create high travel intentions among their followers through their attractiveness, similarity, and expertise [42].

3. Encouraging UGC
Encourage travelers to share their experiences by creating campaigns that highlight user-generated content. This can enhance the destination's image and provide authentic insights for potential travelers.

3.4 Create an Application

This application was developed using Visual Studio Code (VS Code) because it has strong support for the Python programming language. In addition, VS Code allows the development of web-based applications directly with various extensions and features that simplify the coding, debugging process, and integration with web frameworks such as Flask or Django [43, 44].

3.4.1 Create app.py

TABLE 3. App.py script

```
import pickle; import pandas as pd;  import numpy as np
from flask import Flask, request, jsonify, render_template;
from sklearn.preprocessing import StandardScaler
try:
    with open("kmeans_model.pkl", "rb") as model_file:
        kmeans = pickle.load(model_file)
    with open("scaler.pkl", "rb") as scaler_file:
        scaler = pickle.load(scaler_file)
except FileNotFoundError:
    print("Error: Model or scaler file not found. Make sure the
kmeans_model.pkl and scaler.pkl files are in the correct
directory.")
    exit()
app = Flask(__name__)
@app.route("/")
def home():
    return render_template("index.html")
@app.route("/predict", methods=["POST"])
def predict():
    try:
        data = request.get_json()
            if not isinstance(data, list) or not all(isinstance(row,
list) for row in data):
        return jsonify({"error": "Invalid input format.
Expected a list of lists."}), 400
            df = pd.DataFrame(data)
        df_scaled = scaler.transform(df)
        clusters = kmeans.predict(df_scaled)
        return jsonify({"clusters": clusters.tolist()})
    except Exception as e:
        return jsonify({"error": str(e)}), 500
if __name__ == "__main__":
    app.run(debug=True)
```

Table 3 is the form of the app.py script, where this application is used to build a web-based service that functions as a tourist recommendation system. This application utilizes the Flask framework to manage user requests, process tourist data, and present recommendation results based on machine learning algorithms. In addition, this application can receive input from users, process information related to travel preferences, and display suitable destinations in an interactive web interface.After built the function, an interface was created that functions as an interface for inputting data. This file allows users to enter travel preferences, which are then processed by the application—the primary interface for users to input data related to tourist characteristics. Users can quickly enter travel preferences, social media usage patterns, and habits when searching for travel information. The application will process the input data to cluster tourists based on their behavior patterns. This page's interactive and responsive design ensures a comfortable user experience accessing the tourist cluster mapping system.

3.4.2 Appearance

This web-based application displays a form that users must fill out, containing questions about their habits in planning travel trips. This form is designed to collect data regarding social media use, communication methods, and the quality of information used in tourism decision-making. After the user fills in all the questions and submits the form, the system will process the data provided and determine the appropriate tourist cluster based on the user's behavior patterns.

FIGURE 3. Application Interface

Figure 3 shows the results after the user submits the form. After the system processes the input data, the tourist clustering results are displayed at the bottom of the form. In this example, users are categorized as "Conservative Travelers," meaning they are more likely to rely on more traditional travel planning methods than active social media use. This information can help tourism service providers understand the characteristics of tourists and adjust marketing strategies and services that are more targeted.

IV.CONCLUSION

This study proposes a K-Means clustering-based model to segment tourist behavior from social media usage. The results identified three segments—Digital-Savvy, Passive, and Conservative Travelers—each with distinct patterns. These insights provide actionable strategies for the tourism industry to tailor marketing and services. This approach proves effective for analyzing social media-based tourism data and supports more accurate, data-driven decision-making.

REFERENCE

[1] A. Č. Tanković, I. Bilić, and A. Sohor, "Social Networks Influence In Choosing A Tourist Destination," *Journal Of Content Community And Communication*, vol. 15, no. 8, pp. 2–13, Jun. 2022, doi: 10.31620/jccc.06.22/02.

[2] J. Fardous, J. T. Du, P. Hansen, K.-K. R. Choo, and S. (Sam) Huang, "Group trip planning and information seeking behaviours by mobile social media users: A study of tourists in Australia, Bangladesh and China," *J Inf Sci*, vol. 47, no. 3, pp. 323–339, Dec. 2019, doi: 10.1177/0165551519890515.

[3] X. Nan, R. Parthasarathy, and C. Wang, "K-means++ Algorithm Based on Spark Platform in Large-scale Tourism Data Mining in Cloud Computing Environment," in *2024 3rd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC)*, IEEE, Jul. 2024, pp. 290–295. doi: 10.1109/raiic61787.2024.10671331.

[4] Dr. R. Lama, "IMPORTANCE OF SOCIAL MEDIA PLATFORMS IN TOURISM INDUSTRY," in *Futuristic Trends in Management Volume 3 Book 26*, Iterative International Publisher, Selfypage Developers Pvt Ltd, 2024, pp. 107–117. doi: 10.58532/v3bhma26p2ch3.

[5] M. S. Lewis and A. Gui, "The Influence of Social Media on the Travel Planning Process and Tourist Locations," in *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/iciptm59628.2024.10563345.

[6] S. K. Acharjee and T. Ahmed, "The impact of social media on tourists' decision-making process: An empirical study based on Bangladesh," *Journal of Social Sciences and Management Studies*, vol. 3, no. 1, pp. 56–71, Nov. 2023, doi: 10.56556/jssms.v3i1.685.

[7] A. Nag and B. Gilitwala, "A Study of Indians' Buying Behaviour as Influenced by their use of Social Media," *Journal Of Content Community And Communication*, vol. 17, no. 9, pp. 4–15, Jun. 2023, doi: 10.31620/jccc.06.23/02.

[8] S. Khadka and A. K. Khadka, "The Role of Social Media in Determining Tourists' Choices of Nepalese Destinations," *Journal of Logistics, Informatics and Service Science*, vol. 10, no. 3, pp. 180–193, 2023, doi: 10.33168/JLISS.2023.0314.

[9] P. P. Iglesias-Sánchez, M. B. Correia, and C. Jambrino-Maldonado, "The challenge of linking destination online reputation with competitiveness," *Tourism & Management Studies*, vol. 15, no. 1, pp. 35–43, Jan. 2019, doi: 10.18089/tms.2019.150103.

[10] S. K. Deb and N. Mallik, "Effects of Social Media in Tourism Marketing: Outlook on User Generated Content," *Journal of Digital Marketing and Communication*, vol. 3, no. 2, pp. 49–65, Nov. 2023, doi: 10.53623/jdmc.v3i2.316.

[11] R. C. Sonawane and H. D. Patil, "Clustering Techniques and Research Challenges in Machine Learning," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Mar. 2020, pp. 290–293. doi: 10.1109/iccmc48092.2020.iccmc-00054.

[12] R. Lad and P. S. Metkewar, "Review of Machine Learning Classifier Toolbox of Neuroimaging Data," in *2020 IEEE Pune Section International Conference (PuneCon)*, IEEE, Dec. 2020, pp. 175–179. doi: 10.1109/punecon50868.2020.9362360.

[13] S. Velunachiyar and K. Sivakumar, "Some Clustering Methods, Algorithms and their Applications," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 6s, pp. 401–410, Jun. 2023, doi: 10.17762/ijritcc.v11i6s.6946.

[14] B. Wu, "K-means clustering algorithm and Python implementation," in *2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE)*, IEEE, Aug. 2021, pp. 55–59. doi: 10.1109/csaiee54046.2021.9543260.

[15] R. Wyrzykowski, J. Dongarra, E. Deelman, and K. Karczewski, *Parallel Processing and Applied Mathematics*, vol. 10778. in Lecture Notes in Computer Science, vol. 10778. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-78054-2.

[16] J. Somarribas and A. Loteanu, "Using Micro-Processor Vector Instructions to Optimize Unsupervised Machine Learning K-Means Algorithm," in *2020 IEEE 11th Latin American Symposium on Circuits & Systems (LASCAS)*, IEEE, Feb. 2020, pp. 1–4. doi: 10.1109/lascas45839.2020.9068970.

[17] J. Liao, X. Wu, Y. Wu, and J. Shu, "K-NNDP: K-means algorithm based on nearest neighbor density peak optimization and outlier removal," *Knowl Based Syst*, vol. 294, p. 111742, Jun. 2024, doi: 10.1016/j.knosys.2024.111742.

[18] Y. Lian, P. Wang, and G. Guo, "Analysis Model of Efficiency and Accuracy in Big Data Based on Clustering Algorithm K-means," in *2023 International Conference on Internet of Things, Robotics and Distributed Computing (ICIRDC)*, IEEE, Dec. 2023, pp. 433–437. doi: 10.1109/icirdc62824.2023.00085.

[19] M. Gul and M. A. Rehman, "Big data: an optimized approach for cluster initialization," *J Big Data*, vol. 10, no. 1, Jul. 2023, doi: 10.1186/s40537-023-00798-1.

[20] E. Jaballi, S. Gdaim, and N. Liouane, "Design and implementation of FPGA-based hardware acceleration for Machine Learning using OpenCL: A case study on the K-Means algorithm," in *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, IEEE, May 2024, pp. 1–6. doi: 10.1109/iccad60883.2024.10553964.

[21] P. Sinha, L. Dey, P. Mitra, and D. Thomas, "A Hierarchical Clustering Algorithm for Characterizing Social Media Users," in *Companion Proceedings of the Web Conference 2020*, in WWW '20. ACM, Apr. 2020, pp. 353–362. doi: 10.1145/3366424.3383296.

[22] S. S. Pavithra, S. Chitrakala, and C. M. Bhatt, "Spectral Clustering of Events in Social Media Flood Images based on Multimodal Analysis," in *2021 7th International Conference on Electrical Energy Systems (ICEES)*, IEEE, Feb. 2021, pp. 648–653. doi: 10.1109/icees51510.2021.9383727.

[23] N. A. Ghani, S. Hamid, I. A. Targio Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Comput Human Behav*, vol. 101, pp. 417–428, Dec. 2019, doi: 10.1016/j.chb.2018.08.039.

[24] B. Zhurakovskyi, I. Averichev, and I. Shakhmatov, "Using the Latest Methods of Cluster Analysis to Identify Similar Profiles in Leading Social Networks," Kyv, Ukraine: Information Technology and Implementation (IT&I-2023), Nov. 2023.

[25] J. Sara and B. Youssef, "A New Method For Getting The Optimal Number Of Clusters By K-Means Using The Weighted Barycenter," *J Theor Appl Inf Technol*, vol. 15, no. 1, 2024, [Online]. Available: www.jatit.org

[26] E. Hancer, "A new multi-objective differential evolution approach for simultaneous clustering and feature

selection," *Eng Appl Artif Intell*, vol. 87, p. 103307, Jan. 2020, doi: 10.1016/j.engappai.2019.103307.

[27] R. Cassandro, Q. Li, and Z. S. Li, "An Entropy-based Data Reduction Method for Data Preprocessing," in *2023 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, Jun. 2023, pp. 351–356. doi: 10.1109/icphm57936.2023.10194224.

[28] Y. Liu, L. Wang, L. Zhao, and Z. Yu, *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, vol. 1075. in Advances in Intelligent Systems and Computing, vol. 1075. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-32591-6.

[29] N. Saini, K. Sharma, P. K. Sarangi, G. Singh, and L. Rani, "Customer Segmentation using K-Means Clustering," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, Oct. 2022, pp. 1–5. doi: 10.1109/icrito56286.2022.9965147.

[30] M. A. Damos *et al.*, "Enhancing the K-Means Algorithm through a Genetic Algorithm Based on Survey and Social Media Tourism Objectives for Tourism Path Recommendations," *ISPRS Int J Geoinf*, vol. 13, no. 2, p. 40, Jan. 2024, doi: 10.3390/ijgi13020040.

[31] Y. Choe, J. (Jamie) Kim, and D. R. Fesenmaier, "Use of social media across the trip experience: An application of latent transition analysis," *Journal of Travel & Tourism Marketing*, vol. 34, no. 4, pp. 431–443, May 2016, doi: 10.1080/10548408.2016.1182459.

[32] T. Long, "Understanding the Social Media Use in Tourism: An Affordance Perspective," in *Well-Being in the Information Society. Fruits of Respect*, Springer International Publishing, 2020, pp. 169–184. doi: 10.1007/978-3-030-57847-3_12.

[33] T.-Y. Wang and J. Park, "Destination Information Search in Social Media and Travel Intention of Generation Z University Students," *Journal of China Tourism Research*, vol. 19, no. 3, pp. 570–588, Jul. 2022, doi: 10.1080/19388160.2022.2101574.

[34] F. Sharmin, M. T. Sultan, D. Wang, A. Badulescu, and B. Li, "Cultural Dimensions and Social Media Empowerment in Digital Era: Travel-Related Continuance Usage Intention," *Sustainability*, vol. 13, no. 19, p. 10820, Sep. 2021, doi: 10.3390/su131910820.

[35] D. Gupta and Y. Sharma, "Social Media: Its Impact on Youth Travelers and Formation of Sustainable Destination Image," *International Journal of Experimental Research and Review*, vol. 41, no. Spl Vol, pp. 239–249, Jul. 2024, doi: 10.52756/ijerr.2024.v41spl.020.

[36] T. Nabi, A. Joshi, A. Vashisht, M. A. Siddiqui, and R. Dwivedi, "The Impact of Organisational Communication and Transformational Leadership on Employee Green Behaviour and Organisational Effectiveness," *Journal of Content, Community and Communication*, vol. 17, no. 9, pp. 109–130, Jun. 2023, doi: 10.31620/JCCC.06.23/09.

[37] A. Elnur and H. Akgün, "Social media experiences of solo leisure travelers during the travel process: A phenomenological study," *Heliyon*, vol. 11, no. 1, p. e41313, Jan. 2025, doi: 10.1016/j.heliyon.2024.e41313.

[38] T. Oliveira, B. Araujo, and C. Tam, "Why do people share their travel experiences on social media?," *Tour Manag*, vol. 78, p. 104041, Jun. 2020, doi: 10.1016/j.tourman.2019.104041.

[39] F. M. Tsai and T.-D. Bui, "Impact of word of mouth via social media on consumer intention to purchase cruise travel products," *Maritime Policy & Management*, vol. 48, no. 2, pp. 167–183, Mar. 2020, doi: 10.1080/03088839.2020.1735655.

[40] L. Lauvelia, C. Angelica, I. Kusumawardhana, and D. Rosman, "Build Indonesian Gen-Z Trust on Travel Planning Process Through Social Media," in *Opportunities and Risks in AI for Business Development*, Springer Nature Switzerland, 2024, pp. 577–586. doi: 10.1007/978-3-031-65203-5_51.

[41] M. A. Camilleri and M. Kozak, "Interactive engagement through travel and tourism social media groups: A social facilitation theory perspective," *Technol Soc*, vol. 71, p. 102098, Nov. 2022, doi: 10.1016/j.techsoc.2022.102098.

[42] W.-K. Chen, P. H. Silaban, C.-H. Chi, W. E. Hutagalung, and A. D. K. Silalahi, "Using fuzzy sets Qualitative Comparative Configuration to Predict Consumer Travel Behavior from Instagram: The Role of Social Media Influencers," in *2022 International Conference on Information Management and Technology (ICIMTech)*, IEEE, Aug. 2022, pp. 431–436. doi: 10.1109/icimtech55957.2022.9915052.

[43] S. bin Uzayr, *Optimizing Visual Studio Code for Python Development: Developing More Efficient and Effective Programs in Python*. Apress, 2021. doi: 10.1007/978-1-4842-7344-9.

[44] A. Del Sole, *Visual Studio Code Distilled: Evolved Code Editing for Windows, macOS, and Linux*. Apress, 2021. doi: 10.1007/978-1-4842-6901-5.

CANDRA AGUSTINA received her Bachelor's degree in Computer Science from Universitas Dian Nuswantoro, Semarang, in 2004, and her Master's degree in Computer Science from STMIK Nusa Mandiri, Jakarta, in 2012. Currently, she is pursuing her Doctorate degree in Information Systems at Universitas Diponegoro (UNDIP), Semarang, Indonesia. She is a lecturer in the Information Systems Department at Universitas Bina Sarana Informatika, Indonesia. Her research interests include e-tourism, machine learning applications, digital transformation, and information systems development. She can be contacted at email: candra.caa@bsi.ac.id.

EKA RAHMAWATI received her Bachelor's degree in Computer Science from Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jakarta, in 2017, and her Master's degree in Computer Science from the same institution in 2019. She is currently pursuing her Doctorate degree in Information Systems at Universitas Diponegoro, Semarang, Indonesia. She is a lecturer in the Information Systems Department at Universitas Bina Sarana Informatika, Solo, Indonesia. Her research interests include information systems development, machine learning applications, and digital transformation in various industries. She can be contacted at email: eka.eat@bsi.ac.id.