IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5040~5049
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5040-5049

Recommendation of similar X profiles using machine learning algorithms

Ismail Bensassi[1,2], Oussama Ndama[2], Mohamed Kouissi[2], El Mokhtar En-Naimi[2]
[1]Department of Computer Science, Faculty of Sciences, University Mohammed V in Rabat, Rabat, Morocco
[2]Data Science, Artificial Intelligence and Smart Systems (DSAI2S) Research Team, C3S Laboratory, Abdelmalek Essaadi University of Tetuan, Tetuan, Morocco

Article Info

Article history:

Received Oct 2, 2025
Revised Aug 25, 2026
Accepted Sep 20, 2026

Keywords:

K-means clustering
K-nearest neighbors
Machine learning
Metadata-based filtering
Recommendation
Social networks
X profile

ABSTRACT

Identifying similar X (formerly Twitter) profiles is crucial for engagement, personalization, and digital inclusion, especially in contexts with limited data or resources. This article compares two metadata-based similarity methods k-nearest neighbor (KNN) and k-means using only public metadata (followers, verification, and biography). A synthetic ground truth, derived from a balanced partition of the dataset, serves as an evaluation. The results show that KNN achieves 93.7% accuracy, 94.2% precision, 93.1% recall, and 93.6% F1-score, far ahead of k-means (51% in all metrics), thanks to its sensitivity to local similarities in the normalized space. Computation time is under 10 ms per query, with memory usage below 50 MB. Lightweight, reproducible, and independent of interaction logs, this KNN pipeline offers a practical approach for recommendation in data-scarce environments, thereby promoting digital inclusion and equitable access to information.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ismail Bensassi
Department of Computer Science, Faculty of Sciences, University Mohammed V in Rabat
Rabat, Morocco
Email: ismail.bensassi@um5.ac.ma

1. INTRODUCTION

Online social networks, particularly X (formerly Twitter), have transformed how users communicate, share information, and discover digital content, making effective profile recommendation important for improving user satisfaction, engagement, and content visibility [1]. However, the increasing number and heterogeneity of users make it challenging to identify profiles that are structurally similar to a given user. Traditional recommender systems often rely on historical interactions, such as follows, mentions, likes, and retweets. These approaches may be less effective for new or inactive users because of the cold-start problem and limited interaction data. In this context, publicly available profile metadata can provide useful information for estimating user similarity without requiring extensive behavioral histories [2], [3].

This study investigates the recommendation of structurally similar X profiles using numerical attributes, binary indicators, and profile descriptors. The selected features include follower and following counts, verification status, profile-picture availability, and biographical information [2], [3]. These heterogeneous attributes are normalized to facilitate the comparison of users and provide a lightweight representation for profile similarity estimation. Two machine learning approaches, k-nearest neighbors (KNN) and k-means clustering, are compared for identifying similar profiles [4], [5]. KNN focuses on local similarity by identifying the closest profiles to a target user, whereas k-means provides a global representation by grouping users around cluster centroids. This comparison is particularly relevant for heterogeneous social-network profiles that may exhibit different similarity patterns.

Since explicit ground-truth labels for profile similarity are unavailable, a proxy evaluation strategy is adopted based on the assumption that users with similar metadata belong to similar latent groups [6], [7]. Therefore, the results evaluate metadata-based structural similarity rather than direct user preferences or real-world interactions. The experiments assess the predictive performance and computational efficiency of both approaches, particularly in data-scarce and cold-start scenarios [6], [8]. Overall, this work examines whether lightweight and computationally efficient machine-learning methods can provide useful profile recommendations using only publicly available metadata. The comparison between KNN and k-means highlights the suitability and limitations of local-neighborhood and global-clustering approaches and provides a basis for future extensions involving richer textual features, interaction data, and hybrid recommendation strategies.

2. STATE OF THE ART AND PROPOSED APPROACH

Profile and content recommendation is a major research area for improving user engagement and platform scalability. Existing recommender systems mainly follow three paradigms: collaborative filtering (CF), content-based filtering (CBF), and hybrid approaches [9]. CF exploits users' historical interactions and preferences to identify users or items with similar behavioral patterns. However, its effectiveness depends on the availability of sufficient interaction data and may therefore be limited for new or inactive users. CBF, in contrast, relies on the characteristics of users, profiles, or content to identify relevant recommendations based on similarity between their attributes [9]. This approach can be particularly useful when interaction histories are sparse, although its performance depends strongly on the quality and representativeness of the available features. Hybrid approaches combine collaborative and content-based information to benefit from their complementary characteristics and potentially improve recommendation quality. Despite these advantages, conventional approaches may still face challenges related to data sparsity, heterogeneous user profiles, and the cold-start problem. These limitations motivate the investigation of lightweight recommendation methods based on publicly available profile metadata, which can provide useful similarity information without requiring extensive historical interaction data.

2.1. Collaborative filtering

CF exploits historical user–item interactions, such as ratings, clicks, follows, likes, and retweets, to identify users or items with similar behavioral patterns [10]. By analyzing these interactions, CF can generate personalized recommendations based on the preferences and activities of users with similar behaviors. This approach has been successfully applied in many recommendation scenarios because it can capture implicit and explicit user preferences without requiring detailed descriptions of the recommended items. However, its effectiveness generally depends on the availability and quality of historical interaction data. In particular, CF suffers from the cold-start problem because new or inactive users lack sufficient interaction history to establish reliable similarity relationships [11]. Data sparsity can further reduce recommendation quality when only a small proportion of users actively interact with available content. These limitations motivate the use of complementary information, such as publicly available profile metadata, to estimate user similarity when behavioral information is limited or unavailable.

2.2. Content-based filtering

CBF recommends profiles or items based on their intrinsic attributes, such as biography, location, follower-to-following ratio, and other profile characteristics [12]. It can operate without historical interactions, making it suitable for cold-start scenarios [13]. However, its performance and diversity depend on the quality and richness of available metadata [14], [15]. The proposed approach follows this paradigm by relying exclusively on structured and derived metadata.

2.3. Hybrid methods

Hybrid recommender systems combine collaborative and content-based information to improve recommendation accuracy and reliability while reducing the limitations of either approach alone [16]. By integrating users' historical interactions with descriptive information about users or items, hybrid models can exploit both behavioral and content-related similarities. This combination can help alleviate the data sparsity and cold-start problems that affect purely collaborative approaches, particularly when additional profile or content information is available. Hybrid strategies can also provide more robust recommendations by considering complementary sources of evidence rather than relying on a single type of information. However, their effectiveness may depend on the availability, quality, and compatibility of the different information sources. Moreover, combining multiple models can increase system complexity and computational requirements, especially when large-scale social-network data are involved. Therefore, lightweight

approaches based on readily available metadata remain attractive when interaction data are limited or when computational efficiency is an important consideration.

2.4. Machine learning for recommendation

Unsupervised learning is particularly useful when interaction data are sparse or unavailable. K-means groups users according to profile features but may perform poorly with heterogeneous or non-spherical data [17]. KNN identifies locally similar profiles through distance calculations in a normalized feature space, preserving local structure and adapting to each query [5]. Graph-based methods, such as PageRank, hyperlink-induced topic search (HITS), and SimRank, exploit user relationships but require explicit interaction data, limiting their use for new or inactive accounts [18].

2.5. X-specific approaches and research gap

Previous studies have incorporated tweet content using term frequency–inverse document frequency (TF-IDF), Word2Vec, or bidirectional encoder representations from transformers (BERT) [19]. However, these approaches generally require sufficient tweet histories, which may be unavailable for new or inactive accounts. In contrast, this study uses publicly available metadata, including follower count, verification status, biography length, and profile completeness, enabling recommendation under cold-start and data-scarce conditions [3], [20].

Previous hybrid approaches combine metadata and behavioral signals [21], while other studies use profile and behavioral characteristics for user clustering [22]. The approach complements these methods by operating without interaction logs. Despite advances in collaborative, content-based, hybrid, and graph-based recommendation, a gap remains in lightweight frameworks based solely on publicly available metadata. This study addresses this gap through a direct comparison of KNN and k-means for metadata-based profile similarity, with emphasis on computational efficiency and cold-start applicability.

3. METHOD

3.1. Data preparation and feature engineering

The experimental study relies on a dataset of X user profiles provided in the users.csv file, which includes structured metadata such as user ID, display name, follower count, following count, number of tweets, biography, location, verification status, and profile image availability [2]. Given the absence of explicit similarity labels, this study construct a synthetic ground truth by partitioning the dataset into two balanced subsets (group 0 and group 1), under the assumption that users within the same subset exhibit comparable profile characteristics [6].

This synthetic binary ground truth, while simplifying real-world complexity, offers a controlled and reproducible benchmark for comparing algorithmic behavior in the absence of behavioral logs a common scenario in emerging digital ecosystems [7]. This study acknowledge that future work could replace this proxy with semantic validation using verified accounts or interest-based taxonomies. To enable algorithmic processing, this study perform feature engineering to extract three categories of descriptors: i) numerical features: followers_count, friends_count, statuses_count; ii) binary indicators: verified, has_profile_image, has_url; and iii) derived metrics: followers_friends_ratio, description_length.

The selected metadata features capture complementary aspects of profile similarity. Numerical attributes such as follower, following, and tweet counts provide indicators of profile activity and audience size, while binary features such as verification status, profile image availability, and uniform resource locator (URL) presence describe profile characteristics and completeness. Derived features, including the followers-to-following ratio and biography length, provide additional information about audience structure and profile description. By combining these heterogeneous attributes after normalization, the feature representation enables KNN to identify profiles that are locally similar across multiple dimensions rather than relying on a single characteristic. Missing values are imputed using median or mode strategies, and all features are normalized using the StandardScaler from scikit-learn to ensure uniform scale and prevent distance distortion in subsequent similarity computations [4], [23]. Feature normalization is performed using the standard score (z-score), where x is the original feature value, μ is the mean of the feature, and σ is its standard deviation.

$$z = \frac{x-\mu}{\sigma}$$

3.2. K-nearest neighbor for profile recommendation

The KNN algorithm operates by computing pairwise distances typically Euclidean in the normalized feature space to identify the k most similar profiles to a given query [5]. The Euclidean distance between two profile vectors x and y is computed as:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

where x_i and y_i represent the values of the ith feature of the two X profiles, and n denotes the total number of features. Profiles with the smallest Euclidean distances are considered the most similar and are therefore recommended to the target user. For a target user, the system retrieves its k nearest neighbors and treats them as recommended similar accounts.

This study tested values of k ∈ {3, 4, 5, 6, 7} and found peak F1-score at k = 4 (93.6%), with stable performance (±0.3%) for k = 5–6. Larger k values slightly reduced precision due to inclusion of distant neighbors. Hence, fix n_neighbors = 4 in the final model. The KNN implementation uses the Euclidean distance metric, while the algorithm='auto' parameter allows scikit-learn to select an appropriate nearest-neighbor search strategy. This approach leverages local neighborhood structure without requiring global assumptions about data distribution.

3.3. K-means clustering for group-based recommendation

K-means partitions the user set into k disjoint clusters by iteratively minimizing intra-cluster variance [24]. The optimal number of clusters is determined using the elbow method, which evaluates the rate of decrease in inertia (sum of squared distances to centroids) as k increases. For the dataset, the inflection point occurs at k = 4 ~ 5, suggesting a reasonable trade-off between model complexity and cluster cohesion. Once clustering is complete, all members of the same cluster as the target profile are recommended as similar. This method assumes that cluster membership implies semantic or behavioral similarity, though it may struggle with non-spherical or overlapping user distributions.

4. RESULTS AND EVALUATION

The performance of both KNN and k-means was evaluated using an artificially constructed ground truth, in which the users were divided into two balanced groups (group 0 and group 1). The underlying evaluation hypothesis is that a high-quality recommendation should predominantly retrieve profiles belonging to the same group as the query user. Accordingly, the recommendation performance was assessed using accuracy, precision, recall, and F1-score, together with the standard deviation computed over five independent runs. These complementary metrics provide a comprehensive evaluation by considering both correct predictions and different types of classification errors [6], [7]. Using multiple evaluation measures also enables a consistent comparison of KNN and k-means under the same experimental protocol.

4.1. K-nearest neighbor results

4.1.1. Model initialization

The KNN model retrieves the four closest neighbors (n_neighbors = 4) using the 'auto' algorithm. The selected numerical and binary features are normalized to reduce differences in feature scales. The normalized feature space is then used to compute distances between the target and candidate profiles. This configuration preserves meaningful local similarity relationships while ensuring consistency across experiments. The strong performance of KNN can be attributed to its neighborhood-based formulation, which focuses on local proximity in the feature space [4], [10]. By considering the closest profiles, KNN effectively captures fine-grained similarities in heterogeneous user populations [10]. This local approach may therefore provide more suitable recommendations than global clustering methods in the considered setting.

4.1.2. Finding similar profiles

The function named find_similar_profiles, designed to recommend X profiles similar to a target profile using the KNN algorithm accepts multiple inputs: the index of the target user profile, the normalized feature matrix, a pre-trained KNN model, and the number of recommendations to generate. Its core workflow consists of the following steps:

i) Validate the input profile index to ensure it falls within the dataset bounds.
ii) Retrieve the corresponding feature vector for the target profile.
iii) Invoke the KNN model to compute nearest neighbors in the feature space.
iv) Fetch the metadata of the identified similar profiles.
v) Augment the output with two supplementary columns (similarity score and group label).
vi) Return the final list of recommended profiles.

4.1.3. Accuracy calculation

The evaluation of the KNN-based recommendation system yields an accuracy of 93.7%, indicating that 93.7% of the suggested profiles share the same ground-truth group as the query profile. This high

alignment indicates strong consistency in the similarity detection capability of the KNN approach under the adopted synthetic-ground-truth evaluation protocol. Table 1 presents the evaluation metrics for KNN-based profile recommendations on a random subset of X profiles.

Table 1. Performance metrics of the KNN-based recommendation system

Metric	KNN (mean±std) (%)
Accuracy	93.7 ± 0.4
Precision	94.2 ± 0.3
Recall	93.1 ± 0.5
F1-score	93.6 ± 0.4

The results demonstrate that the KNN-based recommendation system achieves high and stable performance across all key classification metrics. Under the adopted evaluation protocol, the proposed KNN approach achieved an F1-score of 93.6%. The model exhibits a strong balance between precision (94.2%) and recall (93.1%), indicating stable performance in identifying profiles assigned to the same synthetic ground-truth group while minimizing both false positives and false negatives. The low standard deviation (<0.5%) across five independent runs further supports the robustness and reproducibility of the approach under the adopted experimental protocol.

4.1.4. Confusion matrix

The confusion matrix evaluates the alignment between the ground-truth group of the target profile and the groups of the recommended profiles, thereby quantifying the number of correct and incorrect predictions. In the KNN-based recommendations, the matrix reveals a high proportion of true positives, underscoring the method's reliability. Specifically: i) group 0: 485 correct predictions and 25 misclassifications and ii) group 1: 452 correct predictions and 38 misclassifications. Figure 1 shows the confusion matrix, underscoring KNN's reliability. This result indicates a high level of agreement with the adopted synthetic ground-truth labels, with an accuracy of 93.7%.

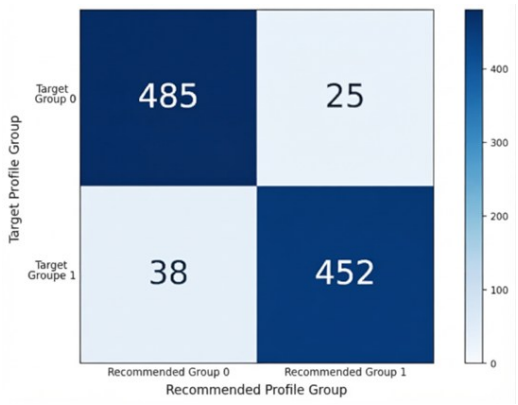

Figure 1. Confusion matrix–similar profile recommendation (KNN)

4.2. Recommendation results using the k-means algorithm

Prior to deploying the k-means algorithm for X profile recommendation, the optimal number of clusters k must be determined. This is achieved using the elbow method, which evaluates the trade-off between model complexity and clustering quality. Specifically, the method plots the inertia defined as the sum of squared distances of samples to their closest cluster centroid against varying values of k. As k increases, inertia monotonically decreases because a larger number of clusters reduces the average intra-cluster distance. The optimal k is selected at the "elbow" point, where the marginal gain in inertia reduction begins to diminish significantly. Figure 2 plots inertia vs. k. The elbow occurs at k = 4 ~ 5. As illustrated in Figure 2, the inertia curve exhibits a pronounced decline for small values of k, followed by a marked flattening of the slope. This inflection point, commonly termed the "elbow," marks the transition where additional clusters yield diminishing returns in variance reduction. For our dataset, this elbow occurs at k = 4 or k = 5, suggesting these values represent an optimal trade-off between clustering fidelity and model complexity.

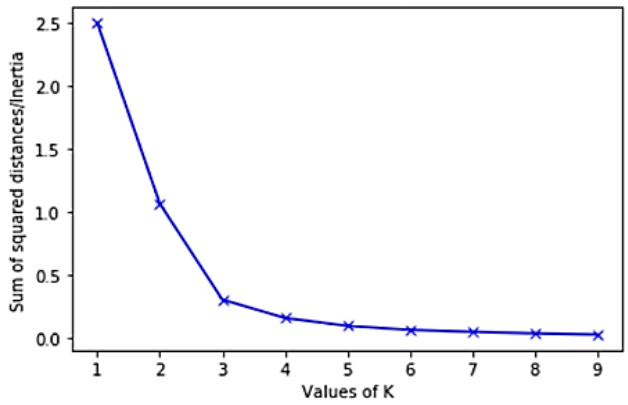

Figure 2. Elbow method for k-means–determining the optimal number of clusters

4.2.1. Searching for similar profiles using k-means

In the context of this article, this function plays a crucial role in implementing the k-means clustering method to recommend similar profiles. by identifying profiles belonging to the same cluster, it allows recommendations to be generated based on the intrinsic similarity of profile characteristics, without requiring explicit interactions between them. Figure 3 shows k-means clusters after principal component analysis (PCA) dimensionality reduction. Overlap between groups is visible, indicating weak separation. The figure shows the k-means clustering results after applying PCA, with the data projected onto a 2D space defined by the first two principal components (PC1 and PC2). The visible overlap between the groups indicates weak cluster separation, suggesting that the global partition produced by k-means does not adequately capture the underlying heterogeneity of the profile metadata.

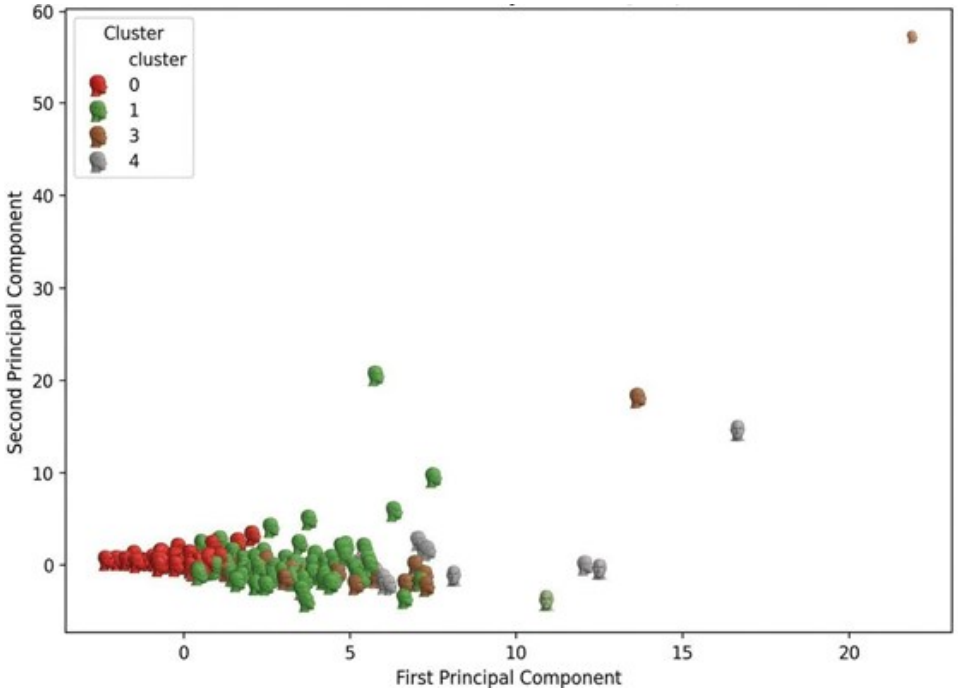

Figure 3. Analysis of clusters found by k-means after dimension reduction with PCA

4.2.2. K-means performance

This study evaluate k-means by treating all intra-cluster profiles as recommendations. Table 2 summarizes the different evaluation metrics of the approach using k-means algorithm of recommendations. The k-means evaluation treats profiles within the same cluster as similar. With an accuracy of 51%, only

about half of the recommendations match the artificial ground truth, indicating that k-means poorly captures profile similarity, likely due to data heterogeneity and non-convex distributions.

Table 2. Performance metrics of the k-means clustering approach for profile recommendation

Metric	K-means (mean ± std) (%)
Accuracy	51.0 ± 1.2
Precision	50.8 ± 1.1
Recall	51.2 ± 1.3
F1-score	51.0 ± 1.2

4.2.3. Confusion matrix

The confusion matrix for k-means enables a comparison between the ground-truth group labels and the cluster assignments generated by the algorithm, revealing how the true classes are distributed across the inferred clusters. Figure 4 reveals a near-uniform distribution across the true groups, providing evidence of poor cluster cohesion. The confusion matrix evaluates the alignment between k-means cluster assignments and the artificial ground-truth labels. It reveals a near-uniform distribution across classes, indicating that the algorithm fails to distinguish meaningful group boundaries which is a sign of poor cluster cohesion and limited discriminative power in this metadata space.

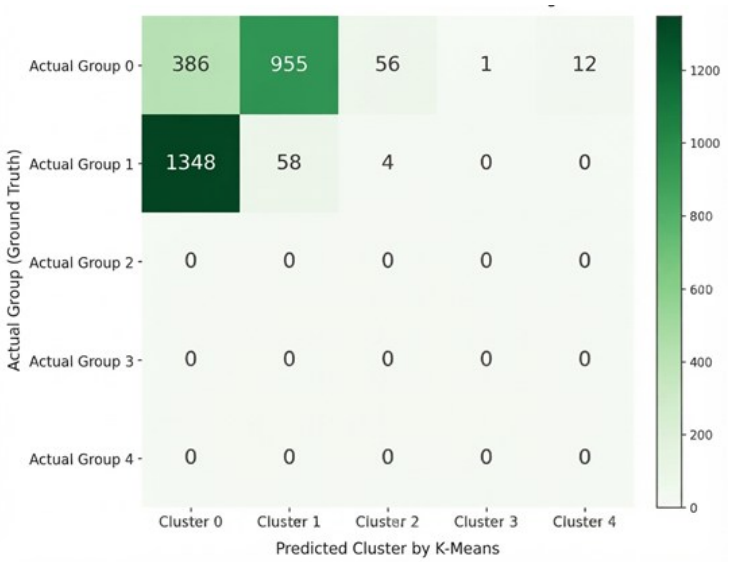

Figure 4. Confusion matrix analysis for k-means clustering

4.3. Statistical robustness and computational evaluation

The results are based on five independent runs and are reported as mean ± standard deviation. KNN showed variability below 0.5 percentage points, supporting the stability of the proposed pipeline under the adopted protocol. Since only aggregated results are available, 95% confidence intervals and formal significance tests cannot be reliably reconstructed. Future evaluations will therefore retain run-level metrics to enable confidence intervals and paired significance testing. The observed query latency below 10 ms and memory usage below 50 MB indicate a lightweight implementation under the tested configuration. However, these results do not establish scalability to larger datasets; future benchmarks should evaluate execution time and memory consumption across different dataset sizes and hardware configurations.

4.4. Robustness to distance measures and alternative clustering baselines

The current KNN implementation uses Euclidean distance after feature standardization. Future experiments will compare it with cosine and Manhattan distances to assess robustness to heterogeneous profile features. K-means could also be compared with density-based spatial clustering of applications with

noise (DBSCAN) and spectral clustering to evaluate whether more flexible clustering structures better capture profile heterogeneity while maintaining computational efficiency.

5. DISCUSSION AND COMPARISON

The results show that KNN substantially outperforms k-means across all reported metrics (93.6% vs. 51% F1-score), consistent with the advantages of instance-based and distance-based similarity methods [5], [23]. KNN preserves local relationships in the normalized feature space, making it suitable for heterogeneous and overlapping social-media profiles. In contrast, k-means imposes a global partition based on centroids and assumes relatively compact clusters, which may not adequately represent heterogeneous or non-convex metadata distributions.

This finding is consistent with Adraoui *et al*. [25], who showed that preserving local structural patterns can improve recommendation quality compared with rigid global clustering. Despite the use of a synthetic binary ground truth, the proposed approach provides a lightweight solution for data-scarce and cold-start scenarios, with no dependency on interaction history [21], [22]. In practice, the KNN module can operate as a lightweight recommendation service: profile metadata are extracted and normalized, nearest profiles are identified, and a ranked list of similar accounts is returned. Because it does not require interaction logs, the module can complement conventional recommenders during cold-start and can be deployed as a low-latency service on resource-constrained infrastructures. The main limitations are the synthetic ground truth, the primary use of Euclidean distance, and the k-means baseline. Future work will evaluate real or expert-validated labels, alternative distances such as cosine and Manhattan, clustering methods such as DBSCAN and spectral clustering, and larger datasets with confidence intervals and significance tests. These extensions would strengthen the validation of robustness and scalability while preserving the lightweight design.

6. CONCLUSION

The significance of this work extends beyond algorithmic accuracy to address real-world challenges in data-scarce digital environments. In regions such as Africa and the Middle East, a growing number of users join social platforms with limited activity history, sparse profile information, or constrained internet access. Conventional recommendation systems dependent on extensive interaction logs often fail under these conditions, thereby widening the digital divide. In response, we propose a lightweight, metadata-only KNN-based recommendation framework that operates effectively without behavioral data. With an F1-score of 93.6% and an observed query latency under 10 ms in the tested configuration, the approach demonstrates strong potential for deployment in low-resource settings, offering a practical and computationally efficient solution with potential for scalable deployment for content discovery, user onboarding, and community formation in critical domains such as education, public health, and civic engagement. Future work will enhance this foundation by incorporating lightweight natural language processing techniques such as TF-IDF weighting or distilled transformer models to extract semantic signals from biographies and tweet content. Additionally, we plan to investigate hybrid architectures that fuse profile metadata with sparse graph-based features (follower networks or mention patterns) to further improve robustness while maintaining low computational overhead. By emphasizing simplicity, reproducibility, and contextual adaptability, this research contributes to the development of inclusive artificial intelligence systems that are not only technically sound but also socially relevant particularly in underserved and infrastructure-constrained regions.

ACKNOWLEDGMENTS

The authors gratefully acknowledge to Data Science, Artificial Intelligence and Smart Systems (DSAI2S) Research Team, C3S Laboratory, Abdelmalek Essaadi University of Tetuan, Morocco for their valuable feedback and technical support during the development of this work.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ismail Bensassi	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Oussama Ndama		✓				✓				✓				
Mohamed Kouissi	✓					✓	✓			✓	✓			
El Mokhtar En-naimi		✓		✓		✓				✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [IB], upon reasonable request.

REFERENCES

[1] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of massive datasets*, 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2014.

[2] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: the million follower fallacy," in *Fourth International AAAI Conference on Weblogs and Social Media*, May 2010, vol. 4, no. 1, pp. 10–17, doi: 10.1609/icwsm.v4i1.14033.

[3] J. Gilbert, S. Hamid, I. A. T. Hashem, N. A. Ghani, and F. F. Boluwatife, "The rise of user profiling in social media: Review, challenges and future direction," *Social Network Analysis and Mining*, vol. 13, Oct. 2023, doi: 10.1007/s13278-023-01146-0.

[4] T. M. Mitchell, *Machine learning*. New York, United States: McGraw-Hill, 1997.

[5] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/bf00153759.

[6] R. B.-Yates and B. R.-Neto, *Modern information retrieval: the concepts and technology behind search*, vol. 48, no. 12. New York, United States: ACM Press, 2011.

[7] S. Tariq, M. Saleem, and M. Shahbaz, "User similarity determination in social networks," *Technologies*, vol. 7, no. 2, 2019, doi: 10.3390/technologies7020036.

[8] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," *25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253–260, 2002, doi: 10.1145/564418.564421.

[9] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, Jun. 2005, doi: 10.1109/TKDE.2005.99.

[10] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *1994 ACM Conference on Computer Supported Cooperative Work (CSCW)*, 1994, pp. 175–186, doi: 10.1145/192844.192905.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *10th International Conference on World Wide Web*, Apr. 2001, pp. 285–295, doi: 10.1145/371920.372071.

[12] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web*, Berlin, Germany: Springer Berlin Heidelberg, 2007, pp. 325–341, doi: 10.1007/978-3-540-72079-9_10.

[13] J. Lu, D. Wu, M. Zhang, A. Zhang, and G. Zhang, "Recommender system application developments: a survey," *Decision Support Systems*, vol. 74, pp. 12–32, Jun. 2015, doi: 10.1016/j.dss.2015.03.008.

[14] P. D. Meo, A. Nocera, G. Terracina, and D. Ursino, "Recommendation of similar users, resources and social networks in a social internetworking scenario," *Information Sciences*, vol. 181, pp. 1285–1305, doi: 10.1016/j.ins.2010.12.001.

[15] J. Dang, M. Bi, J. Yang, and Y. Sun, "Deep learning-based framework for social media user behavior analysis and prediction," *Journal of Circuits, Systems, and Computers*, vol. 34, no. 14, 2025, doi: 10.1142/S0218126625501543.

[16] F. Cacheda, V. Carneiro, D. Fernández, and V. Formoso, "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems," *ACM Transactions on the Web*, vol. 5, no. 1, 2011, doi: 10.1145/1921591.1921593.

[17] P. Han, J. Hong, A. Rasool, H. Chen, Y. Pan, and Q. Jiang, "A hybrid recommendation model for social network services using Twitter data," in *Web Services – ICWS 2022*, Cham, Switzerland: Springer, 2022, pp. 122–129, doi: 10.1007/978-3-031-23579-5_9.

[18] D. L.-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, May 2007, doi: 10.1002/asi.20591.

[19] X. Cheng, X. Yan, Y. Lan, and J. Guo, "BTM: topic modeling over short texts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, Dec. 2014, doi: 10.1109/TKDE.2014.2313872.

[20] S. B. Dhaou, M. Kharoune, A. Martin, and B. B. Yaghlane, "Belief approach for social networks," in *Belief functions: theory and applications*, Cham, Switzerland: Springer, 2014, pp. 115–123, doi: 10.1007/978-3-319-11191-9_13.

[21] G. Mohamed, H. R. Mohamed, T. Adil, M. Moustafa, "Enhancing Twitter sentiment classification with a hybrid bio-inspired feature selection approach," *Journal of Information Systems Engineering and Management*, vol. 10, no. 53s, pp. 74–89, 2025, doi: 10.52783/jisem.v10i53s.10840.
[22] A. Naik, H. Maeda, V. Kanojia, and S. Fujita, "Scalable Twitter user clustering approach boosted by personalized PageRank," in *Advances in Knowledge Discovery and Data Mining*, Cham, Switzerland: Springer, 2017, pp. 472–485, doi: 10.1007/978-3-319-57454-7_37.
[23] S. Alagu and R. Jayasuji, "Distance metric optimization and comparative evaluation in machine learning frameworks," *Zhuzao/Foundry*, vol. 28, no. 12, pp. 36–44, 2025.
[24] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
[25] M. Adraoui, S. Souabi, A. Retbi, M. K. Idrissi, and S. Bennani, "Towards a hybrid recommendation approach using a community detection and evaluation algorithm," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6718–6728, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6718-6728.

BIOGRAPHIES OF AUTHORS



Ismail Bensassi is an associate professor at the Department of Computer Science, Faculty of Science, Mohamed V University, Rabat. He is member of the ANISSE Research Team, Department of Computer Science, Faculty of Sciences, University Mohammed V in Rabat, Morocco and E-DSAI2S (Data Science, Artificial Intelligence and Smart Systems Research Team), Faculty of Sciences and Technologies, Abdelmalek Essaadi University of Tetuan, Morocco. He holds a Ph.D. in Artificial Intelligence and Big Data and Computer Science Engineer, Laureate of Faculty of Sciences and Techniques of Tangier. The research topics of interest are smart systems, sentiment analysis, smart cities, multi-agent systems, CBR, ontology, machine learning, deep learning, and big data. He can be contacted at email: ismail.bensassi@um5.ac.ma.



Oussama Ndama is a Ph.D. student in DSAI2S (Data Science, Artificial Intelligence and Smart Systems Research Team), C3S Laboratory, Faculty of Sciences and Technologies, Laureate of Tetuan, Morocco. He had his master in Computer Science and Big Data, Laureate of FST of Tangier. He is also a business intelligence engineer with more than 5 years of experience in different multinational companies. The research topics of interest are smart systems, machine learning, deep learning, natural language processing, ANN, sentiment analysis, and smart cities. He can be contacted at email: oussama.ndama@etu.uae.ac.ma.



Mohamed Kouissi is an associate professor at the Polydisciplinary Faculty of Larache, Abdelmalek Essaadi University. He is member in DSAI2S (Data Science, Artificial Intelligence and Smart Systems Research Team), Faculty of Sciences and Technologies, Abdelmalek Essaadi University of Tetuan, Tangier, Morocco. He is an engineer in Computer Science and specialized in software mobile development. Laureate of Tangier National School of Applied Sciences in 2011. The research topics of interest are MAS, CBR, machine learning, smart cities, and e-learning. He can be contacted at email: m.kouissi@uae.ac.ma.



El Mokhtar En-Naimi is a full professor at the Abdelmalek Essaadi University, Faculty of Sciences and Technologies of Tangier, Morocco. He served as head of the Department of Computer Sciences from 2016 to 2020 and was responsible for the LST Computer Engineering program from 2012 to 2016. Since 2022/2023, he has led the Data Science, Artificial Intelligence and Smart Systems (DSAI2S) Research Team, Faculty of Sciences and Technologies, Abdelmalek Essaadi University of Tetuan, Tangier, Morocco. He is also a founding member of the C3S (Computer Science and Smart Systems) Laboratory. He has served as an expert evaluator with ANEAQ and the Abdelmalek Essaadi University. His research interests include artificial intelligence, machine learning, deep learning, data mining, MAS, e-learning, big data, wireless sensor networks, VANET/MANET, and smart cities. He has authored and co-authored numerous international publications and supervised several doctoral theses in Computer Science. He has also served as a chair, committee member, session chair, and reviewer for international conferences and workshops. Since 2009, he has been an associate member of the Institute of Complex Systems in Normandy, University of Le Havre, France. He can be contacted at email: en-naimi@uae.ac.ma.