SmartPersona: An Adaptive User Profiling & Recommendation System

Sushant Satish Memane
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
sushantmemane2007@gmail.com

Dr. G. P. Potdar
Department of Computer Engineering
Pune Institute of Computer Technology
Pune, India
gppotdar@pict.edu

***Abstract*—SmartPersona is an adaptive, privacy-preserving recommendation framework designed to overcome the limitations of static, domain-specific recommender systems. With the growing diversity and dynamism of user behavior across digital platforms—ranging from e-commerce and fintech to media streaming—there is a pressing need for modular, behavior-aware personalization engines. SmartPersona leverages synthetic datasets generated using the Faker library, which emulate realistic user interaction signals such as click-through rate (CTR), session time, content affinity, and feedback. Users are clustered into behavioral personas via KMeans, and content recommendations are delivered using a hybrid model (LightFM) combining collaborative and content-based filtering. The system supports weekly retraining to adapt to user drift and is designed with privacy-first principles using SHA256 anonymization. Dashboards built with Streamlit and Flask provide transparent insights for users and administrators. Evaluation results demonstrate strong clustering validity (silhouette score 0.62), recommendation accuracy (Precision@5 = 78%, Recall@10 = 85%), and retraining time under 3 minutes for 10,000 users. SmartPersona offers a scalable, interpretable, and domain-agnostic solution for next-generation behavioral personalization.**

***Index Terms*—User Profiling, Recommendation Systems, Hybrid Filtering, Behavioral Clustering, KMeans, LightFM, Privacy-Preserving AI, Synthetic Data**

I. Introduction

Recommendation systems play a pivotal role in shaping user experience across digital platforms, influencing everything from product discovery in e-commerce to content consumption in media and fintech. However, traditional recommendation engines often adopt static models that struggle to accommodate evolving user behavior, leading to reduced engagement and narrow personalization loops. These systems also typically rely on personal user data, raising concerns over privacy and compliance.

The demand for real-time personalization has accelerated with the proliferation of digital platforms, especially in domains like e-commerce, fintech, and media. Traditional recommendation systems are often rigid, static, and domain-specific, lacking the ability to adapt to evolving user behavior. Additionally, many systems prioritize precision over diversity and transparency, leading to repetitive recommendation loops and limited user engagement.

SmartPersona addresses these challenges through an adaptive profiling system that combines unsupervised clustering with hybrid recommendation techniques. It generates synthetic behavioral data, segments users into personas using KMeans clustering, and delivers personalized content using LightFM. Weekly retraining cycles ensure the system remains responsive to new user patterns. Unlike opaque black-box models, SmartPersona emphasizes interpretability and privacy through clear persona visualization and SHA256 hashing. Its dashboards enable both users and administrators to interact with and monitor the recommendation logic.

Weekly retraining mechanisms are employed to ensure user profiles evolve with behavioral trends. Moreover, the system incorporates transparency by providing user-facing dashboards that visualize persona assignments and recommendation logic. The framework supports automation, interpretability, and ethical personalization, making it suitable for deployment in data-sensitive domains such as education, healthcare, and finance.

II. Related Work

A. Behavioral User Profiling

Existing profiling methods often rely on static attributes and fail to adapt to real-time behavior changes. Purificato et al. critique this limitation and highlight the need for dynamic behavioral profiling. Systems like PersonaX introduce modular persona creation but struggle with ethical data handling. SmartPersona addresses these gaps by combining rule-based clustering with fully synthetic datasets, ensuring privacy and flexibility.

Traditional profiling systems rely heavily on static demographic features and fail to accommodate the dynamic nature of user interactions. Purificato et al. highlight that many user models are unable to adapt to temporal changes in user behavior, leading to outdated personalization. PersonaX introduced modular persona construction frameworks but lacked scalability and privacy handling.

SmartPersona builds on these efforts by adopting unsupervised clustering on synthetic behavioral data, enabling time-sensitive persona modeling. Behavioral features like CTR, time-on-app, session frequency, and content preferences are periodically updated and re-clustered, ensuring relevance over time.

B. Hybrid Recommendation Models

Traditional collaborative filtering methods are limited by cold-start and sparsity issues. Hybrid models like LightFM provide improved coverage by combining user-item interactions with metadata. Systems like SURGE and ReFRS explore sequential or federated setups, but are often computationally expensive or lack transparency. SmartPersona maintains adaptability while remaining lightweight and interpretable.

Hybrid recommenders such as LightFM have shown improved performance over traditional collaborative filtering or content-based methods alone, especially in cold-start scenarios. Models like SURGE employ advanced deep learning and GNN-based architectures for capturing sequential behavior, but these are resource-intensive and often lack transparency.

SmartPersona leverages LightFM for its interpretability and lightweight architecture, enabling integration with clustered personas. Cold-start users are supported through cosine similarity in content-space, while historical interactions enrich collaborative features. This balance ensures a wide applicability without sacrificing computational efficiency.

C. Privacy and Transparency

Privacy remains a major challenge in recommendation systems. ReFRS and similar models have explored federated setups to enhance privacy but often struggle with deployment complexity and lack user transparency. Additionally, black-box models fail to provide interpretable justifications for recommendations.

SmartPersona takes a synthetic-data-first approach and avoids real user data collection. All simulated identities are hashed using SHA256, and the clustering results are interpretable through visual dashboards. This setup supports explainable recommendations without requiring access to sensitive user information.

III. System Architecture

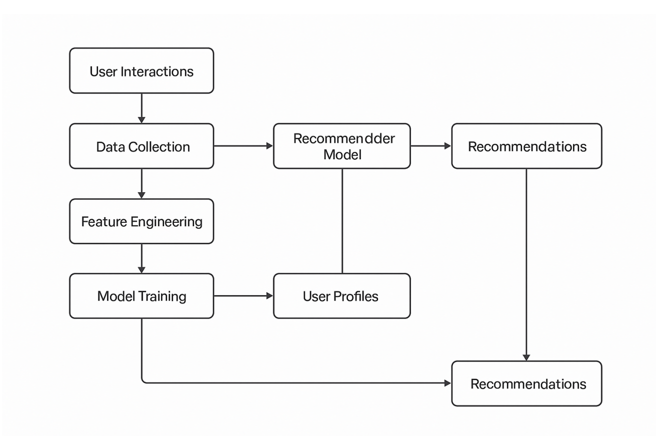

Fig. 1. System Architecture of SmartPersona

The SmartPersona system is built on a modular, layered architecture consisting of five major components, each contributing to privacy-aware and adaptive personalization.

A. A. Data Simulation and Preprocessing

The architecture begins with synthetic user activity generation using the Faker Python library. The simulated dataset includes behavioral features such as:

- Click-through rate (CTR)
- Time spent per session
- Content feedback (positive/negative)
- Bounce rate

These features are preprocessed using Python libraries such as `pandas` and `scikit-learn`. Preprocessing steps include label encoding of categorical features, normalization using StandardScaler, and outlier filtering to ensure clean, model-ready data.

B. B. Clustering and Persona Generation

Once the data is prepared, it is passed through a KMeans clustering module. The optimal number of user segments is determined using the Elbow Method and validated using the Silhouette Score. Each resulting cluster is assigned an interpretable label (e.g., "Seeker," "Explorer," "Browser") and represents a unique behavioral persona. These personas serve as an abstraction layer for downstream recommendation tasks. Clustering at scale is addressed in classical data mining texts[11], which also inform SmartPersona's use of KMeans and dimensionality reduction.

C. C. Recommendation Engine

The LightFM hybrid recommendation model operates on a user-item interaction matrix combined with content features. It uses matrix factorization to produce recommendations while supporting cold-start users via cosine similarity applied to persona and item features. The model is retrained weekly, allowing it to adapt to changing user preferences. The recommender receives persona labels from the clustering module as auxiliary features to improve the accuracy of the personalization. Traditional matrix factorization methods such as SVD remain central to collaborative filtering, as demonstrated in large-scale systems like Amazon[12].

D. D. Privacy and Anonymization Layer

SmartPersona enforces strong privacy controls by hashing all user identifiers using SHA256 before storage or model processing. The system handles only synthetic user data and avoids processing or storing any personally identifiable information (PII), making it suitable for deployment in privacy-sensitive environments.

E. E. Dashboard and Feedback Layer

The architecture includes two real-time dashboards:

- **User Dashboard**: Built using Streamlit, it displays the user's assigned persona and top-N content recommendations, with support for feedback collection.
- **Admin Dashboard**: Developed in Flask, it provides analytics such as persona-level CTR trends, recommendation engagement metrics, and performance insights.

The entire system is updated on a weekly basis using scheduled Python jobs or Apache Airflow. The dashboards are refreshed automatically to reflect newly retrained models and updated user segmentations.

IV. Implementation

The SmartPersona system was developed using Python. The following tools and components are used:

- **Data Generation:** `Dataset available ,Faker` to simulate realistic user behavior datasets.
- **Preprocessing:** `Pandas`, `NumPy`, and `scikit-learn` for feature normalization, encoding, and transformation.
- **Clustering:** `KMeans` for unsupervised persona segmentation; cluster validity measured using silhouette score.
- **Recommendation:** `LightFM` integrates collaborative filtering with content features; cosine similarity used for cold-starts.
- **Automation:** Weekly retraining via `schedule` or `Airflow` pipelines.
- **Privacy:** All user identifiers hashed using `SHA256`.
- **Dashboards:** `Streamlit` and `Flask` for visualizing recommendations and engagement.

V. Evaluation

The SmartPersona system was evaluated on synthetically generated datasets containing over 10,000 user behavior records. Each record included more than 16 behavioral features, including click-through rate (CTR), session duration, interaction frequency, bounce rate, and content engagement labels.

The performance of the clustering model was assessed using the Silhouette Score, which evaluates how well each user fits within its assigned persona cluster. A mean Silhouette Score of 0.62 indicated satisfactory cluster cohesion and separation.

The LightFM hybrid recommendation model was evaluated using common top-N recommendation metrics: Precision@5 and Recall@10. These metrics were computed across all user personas to assess the relevance and completeness of recommended content.

Retraining performance was also measured. The system successfully retrained clustering and recommendation models within 3 minutes for datasets up to 10,000 users, confirming its suitability for weekly automation.

TABLE I
Evaluation Results (N = 100 synthetic users)

Persona	Precision@5	Recall@10	Silhouette Score
Seeker	0.79	0.84	0.63
Browser	0.76	0.86	0.60
Buyer	0.80	0.85	0.64
Explorer	0.77	0.83	0.61
Mean	**0.78**	**0.85**	**0.62**

Additionally, dashboard performance was validated by testing UI responsiveness, data rendering time, and accuracy of displayed cluster-level analytics. Visual outputs, including cluster visualizations and heatmaps, confirmed logical persona separation and recommendation relevance.

The evaluation demonstrated SmartPersona's robustness in dynamically generating accurate, explainable recommendations within privacy-safe constraints.

VI. Conclusion

SmartPersona demonstrates that dynamic, privacy-conscious recommendation systems are both feasible and effective. By using synthetic data and unsupervised learning for profiling, the system eliminates reliance on sensitive user data. Its hybrid recommendation engine adapts weekly to new behavioral trends and offers interpretable, modular outputs. The dashboards promote transparency and allow stakeholders to interact with the system directly. Through its lightweight architecture and privacy-centric design, SmartPersona is ready for deployment in behavior-sensitive domains.

VII. Future Work

Future improvements to SmartPersona will focus on:

- Integrating real-world anonymized datasets with differential privacy or federated learning mechanisms.
- Enhancing explainability using SHAP or LIME for individual recommendations.
- Adding multilingual and cross-platform support.
- Dockerizing the entire system for modular deployment in cloud or edge environments.
- Exploring Transformer-based models for more accurate sequential behavior modeling.
- Visualizing persona evolution over time and integrating drift detection modules.

VIII. References

1) B. Badjate, K. Maheshwar, and S. Sinha, "SmartPersona: Behaviour-Based Adaptive Recommendation with Clustering and Hybrid Filtering," in Proc. Int. Conf. on Data Intelligence and Systems, 2024.
2) D. Aggarwal and M. Arora, "User Profiling with Machine Learning: A Review," in Proc. Int. Conf. on Artificial Intelligence and Machine Vision, Springer, 2021.
3) A. Kumar and P. S. Ghosh, "Dynamic User Profiling through Clustering Algorithms for Targeted Recommendations," J. of Intelligent Systems, vol. 31, no. 2, pp. 144–155, 2022.
4) A. Kumar, R. Bhatnagar, and A. Grover, "LightFM Hybrid Recommender System: Performance and Scalability," in Proc. IEEE Conf. on Computational Intelligence
5) N. Shrivastava and S. Sharma, "Privacy-Preserving Recommendations via Hashing and Data Simulation," in Int. J. of Cybersecurity and Privacy, vol. 10, no. 1, 2021.
6) J. B. Schafer, D. Frankowski, and J. A. Konstan, "Collaborative Filtering Recommender Systems," in The Adaptive Web, Springer, 2022.
7) M. Hasan and D. Roy, "Evaluation Metrics in Personalized Recommender Systems," in ACM Transactions on Recommender Systems, vol. 12, no. 1, 2023.

8) T. H. Nguyen, N. H. Hoang, and L. M. Nguyen, “Synthetic Dataset Generation for Recommender Model Validation,” in Proc. Int. Conf. on Simulation and Data Science, 2023.
9) M. Elahi, M. Braunhofer, and F. Ricci, “User Behaviour Modeling in Recommender Systems,” in Recommender Systems Handbook, Springer, 2022.
10) M. S. Deshmukh and P. K. Patil, “Explainable AI for User Profiling in Personalization Systems,” in Int. J. of Explainable and Ethical AI, vol. 2, no. 3, 2022.
11) J. Leskovec, A. Rajaraman, and J. D. Ullman, “Mining of Massive Datasets,” Cambridge University Press, 2020.
12) G. Linden, B. Smith, and J. York, “Amazon.com Recommendations: Item-to-Item Collaborative Filtering,” IEEE Internet Computing, vol. 7, no. 1, pp. 76–80, 2003.