Title: Interest-based Content Recommendation — Approach & Evaluation

---

Problem statement:

Recommend top-3 posts for each user using profile interests, past engagement, and post attributes.

Problem Understanding:

The task is to recommend **top 3 posts for each user** based on profile interests, past engagements, and content attributes. Three datasets were provided:

- **Users.csv** → user profile information (e.g., interests).
- **Posts.csv** → post metadata (topics, attributes).
- **Engagements.csv** → interaction logs (views, likes, comments, shares, etc.).

The goal is to combine **content-based filtering** (user–post similarity) with **collaborative filtering** (implicit feedback from engagement data).

Approach

1. Data Preparation:

- Loaded and merged the three datasets.
- Standardized column names for consistency.
- Assigned **weights to engagement types** (view=1, click=1, like=3, comment=5, share=4) to capture interaction strength.

2. Content-Based Filtering:

- Represented posts and users in a TF-IDF vector space using interests and attributes.
- Computed **cosine similarity** between users and posts to score relevance.

3. Collaborative Filtering:

- Built a **user–item matrix** from engagement data with weighted scores.
- Applied **Truncated SVD** (latent factor model) to extract hidden patterns in user–post interactions.
- Generated post recommendations by ranking posts based on latent similarity.

4. Hybrid Model:

- Combined **content similarity score** and **collaborative score** using a weighted sum (default α=0.5).
- For each user, ranked all posts and extracted **top 3 recommendations**.
- Exported results to recommendations_top3.csv.

Metrics

Since ground truth relevance was not explicitly provided, evaluation is based on:

- **Coverage**: Ensures each user gets recommendations.
- **Diversity**: Recommendations span multiple interests/topics.
- **Engagement alignment**: Higher weight given to posts with strong engagement signals (like, share, comment).

If labelled relevance data were available, standard metrics like **Precision@3**, **Recall@3**, or **NDCG** would be used.

Possible Extensions

1. **Hyperparameter Tuning:**
    - Adjust number of SVD components, similarity weighting (α), and engagement weights.
    - Incorporate more contextual features (time decay, recency).
2. **Advanced Models:**
    - Use **Neural Collaborative Filtering (NCF)** or **Matrix Factorization with implicit feedback**.
    - Graph-based recommendation using user–post interaction networks.
3. **Personalization Enhancements:**
    - Time-aware models to prioritize recent interests.
    - Re-ranking based on diversity and novelty to avoid echo chambers.
4. **Evaluation Improvements:**
    - If engagement logs are split into train/test, compute **offline metrics**.
    - Deploy online A/B testing for real-world performance validation.

Final Output

- **Top 3 recommended posts for each user** are stored in:

  recommendations_top3.csv
- Documentation provided in both **PDF** and **DOCX** formats for clarity.

Example Rows from recommendations_top3.csv

user_id	recommended_posts
U1	[P1, P88, P57]
U2	[P46, P51, P96]
U3	[P45, P58, P81]
U4	[P58, P39, P50]
U5	[P46, P79, P37]

(Here, Uxxx represents users and Pxxx represents posts.)
