Recap

21 January 2025 15:56

Phase I → Content Based → Attributes of song

Phase II → Collaborative filtering → Item – User recommendations

Phase III → Hybrid → Weighted approach

Content Based → Music Info → 50k

↓

transform → input vector

Similarity scores

Collaborative filtering

30k

User Id track Id Playcount

Music Info → (30k) → filtered data

Sparse matrix

30k, 1million

User ids

track ids → 0 → Playcount → 60 GB.

↓

Scipy sparse matrix → 30 MB

integer encoded

ind → track ids.

30k → trackId → Interactionmatrix

→ track ids.

30k → trackId → Interaction matrix

↓

input vector → Sort ← track ids ← top songs information

Hybrid Recommender system → Weighted approach

↓

dynamic

Deployment → DVC Pipeline

CI → GitHub Runner → code + DVC Pull

Test → Pytest 200

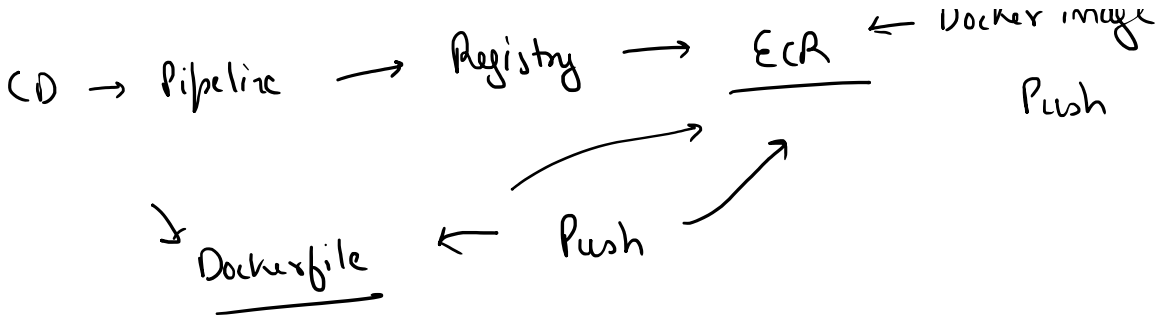
CD → Docker image → local

↓

Dockerfile → instructions

Run → test manually

CD → Pipeline → Registry → ECR ← Docker image Push


CD → Pipeline → Registry → ECR
← Docker image
Push
Dockerfile
← Push

Docker

21 January 2025 15:56

Dockerfile → line → instruction

↓

layer in image

Install Python

FROM Base Image → Python ✓

WORKDIR ← ✓

Dockers layers.

COPY requirements.txt ✓

RUN pip install -r reqs.txt → Static

COPY → datasets ✓ Caching

COPY → Python.py ✓ traffic

EXPOSE → 8000

CMD → Streamlit run

Mlflow → Docker tags for versioning

Mlflow → Docker tags for versioning