# Pipeline Architecture and Documentation

## Pipeline Steps

1) Ingest raw CSVs.

2) Clean & impute missing values.

3) Feature engineering (rolling features, liquidity\_ratio).

4) Train models with time-aware split.

5) Evaluate and save the best model.

6) Deploy via Streamlit/Flask.

## Data Flow

Data flows from raw CSV -> cleaned CSV -> feature-enriched CSV -> model training -> saved model -> deployment UI/API.

## Operational Notes

Run scripts in order: data\_preprocessing.py -> feature\_engineering.py -> eda.py -> model\_training.py -> app.py

Ensure /mnt/data paths exist and have read/write permissions.