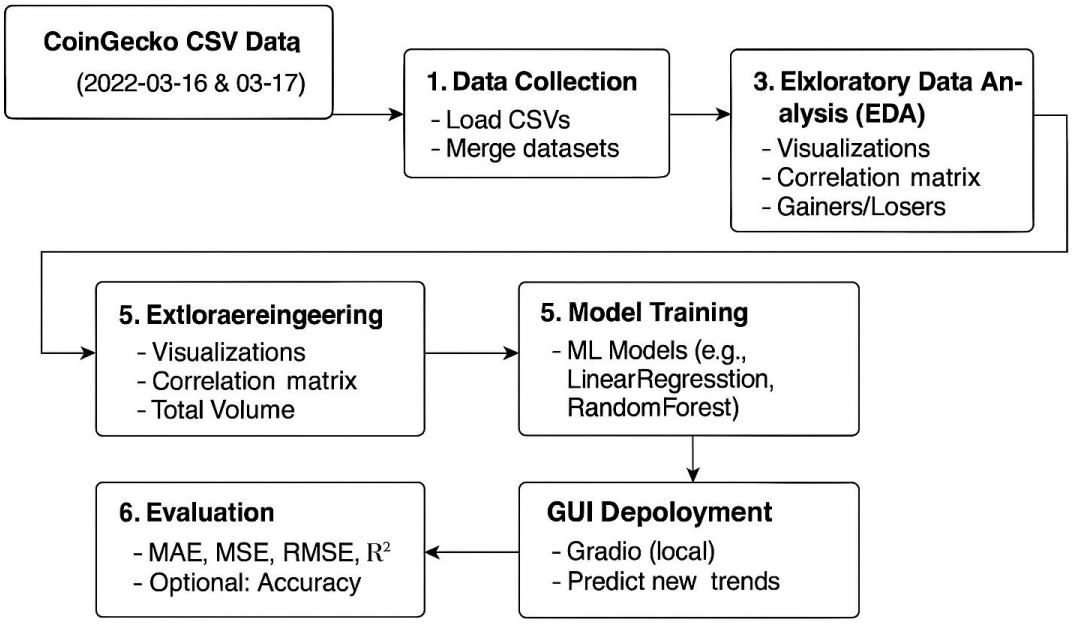
Pipeline Architecture & Document

➢ Architecture:



➢ Document:

1. Objective

To build a machine learning pipeline that analyzes historical cryptocurrency data and predicts price-related trends or metrics.

2. Pipeline Stages

2.1 Data Collection

- Source: CoinGecko CSVs:
("coin_gecko_2022-03-16.csv")
("coin_gecko_2022-03-17.csv")
- Method: Use pandas to read and merge daily data

2.2 Data Preprocessing

- Handle missing values and duplicates
- Convert data types
- Normalize or scale values

2.3 Exploratory Data Analysis (EDA)

- Plot price changes, rank distributions
- Analyze correlations between features
    Distribution plots
    Correlation matrix
    Top gainers/losers by % change

2.4 Feature Engineering

- Create new features such as:
  - price_change_percentage_24h
  - market_cap_rank
  - total_volume

2.5 Model Training

- Used algorithms
- Apply train_test_split to divide the data

2.6 Model Validation

- Evaluate using:
  - Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)
- R^2 Score

2.7 Hyperparameter Tuning

- Optimize model using:
  - Grid Search or Randomized Search
  - Cross-validation (CV)

2.8 Inference / Deployment

- GUI Deployment by gradio framework (locally)
- Create prediction functions for real-time input