

Phase 2 – Preprocessing, Feature Engineering, and Storage

Cleaning & Standardization

- Removed nulls and fixed column inconsistencies
- Standardized time indexes and fixed timezone errors
- Renamed confusing columns like "Unnamed: 4" to "Close"

Feature Engineering

- Added RSI, EMA, MACD indicators to static datasets
- Calculated rolling averages and volatility zones

Dataset Preparation

- Saved all enriched files into `data/processed/enriched/`
- Scaled closing prices using MinMaxScaler
- Generated 60-step sequences for training

Output

- Stored processed model-ready datasets as `.npz` files
- File naming organized by ticker and source
- Separate storage for static, live, and global categories