ML Trading Bot – Data Dictionary & Metrics Guide

# 1) Skąd bierzemy dane i jak wyglądają pliki

Pipeline:  
  
1. download\_data.py – pobiera surowe dane OHLCV.  
2. tech\_indicators.py – liczy wskaźniki techniczne.  
3. runner\_batch – generuje sygnały i metryki.  
4. score\_strategies – wybiera TOP strategie.  
5. ensemble – buduje sygnały zespołowe.  
6. prepare\_full\_ensemble.py – łączy sygnał z cenami.  
7. score\_ensemble.py – liczy metryki dla ensemble.  
8. api.main + dashboard.py – serwują i wizualizują wyniki.

# 2) Surowe pola rynkowe (OHLCV)

- Date – znacznik czasu  
- Open – cena otwarcia  
- High – najwyższa cena  
- Low – najniższa cena  
- Close – cena zamknięcia  
- Adj Close – cena skorygowana  
- Volume – wolumen obrotu

# 3) Wskaźniki techniczne (feature store – processed)

- SMA – Simple Moving Average (np. SMA\_20)  
- EMA – Exponential Moving Average (EMA\_20)  
- RSI – Relative Strength Index (RSI\_14)  
- MACD – Moving Average Convergence Divergence (MACD\_12\_26\_9)  
- Bollinger Bands – BBL, BBM, BBU, BBP  
- STOCH – Stochastic Oscillator (STOCHk\_14\_3\_3, STOCHd\_14\_3\_3)

# 4) Sygnały transakcyjne (signal)

Wartości:  
- 1 – kup (long)  
- 0 – nic (flat)  
- -1 – sprzedaj  
  
Strategie: macd, rsi, sma, ema, bollinger

# 5) Metryki backtestu

- Sharpe Ratio  
- Sortino Ratio  
- Max Drawdown (MDD)  
- Win Rate  
- CAGR (Compounded Annual Growth Rate)  
- MAR = CAGR / |MDD|  
- Final Equity

# 6) Czynniki annualizacji

- 1h – ~1638  
- 1d – 252  
- 1wk – 52

# 7) Konwencja nazw plików

- raw: {TICKER}\_{INTERVAL}\_{DATE}.csv  
- processed: ...\_indicators.csv  
- backtest: ...\_indicators\_{strategy}\_backtest.csv  
- wyniki: batch\_results.csv, ensemble\_batch\_results.csv

# 8) Pliki wynikowe i rankingi

- runner\_batch → batch\_results.csv  
- score\_strategies → top\_overall.csv, top\_per\_bucket.csv  
- ensemble → ensemble\_full.csv

# 9) Ensemble – jak składamy strategie

- Wybór TOP z top\_per\_bucket.csv  
- Głosowanie (mean > 0.5)  
- Zapisywanie jako ensemble.csv i ensemble\_full.csv

# 10) API – najważniejsze endpointy

- GET /tickers, /intervals, /strategies  
- GET /results  
- GET /signals  
- POST /signals/aggregate  
- POST /equity/aggregate

# 11) Dashboard (Streamlit)

- Lista tickerów/interwałów/strategii  
- Agregacja sygnałów  
- Metryki i wykresy equity curve

# 12) Konfiguracja config.yaml

project:  
 timezone: Europe/Warsaw  
data:  
 tickers: [AAPL, MSFT, TSLA]  
 intervals: [1h, 1d, 1wk]  
backtest:  
 initial\_cash: 100000  
 strategies: [macd, rsi, sma, ema, bollinger]  
scoring:  
 top\_n: 20  
 metric: sharpe

# 13) Dodatkowe pojęcia

- Equity curve  
- drawdown series  
- signal lag  
- rebalancing rules