ML Model Journal

James Evans

March 2025

1 Introduction

2 Feature Engineering

The feature engineering process plays a crucial role in transforming raw stock market data into meaningful indicators that can be used for machine learning models. In this section, we provide a detailed explanation of the calculations implemented in `feature_engineering.py` and how they contribute to the overall predictive modeling pipeline.

2.1 Overview

The purpose of feature engineering is to extract relevant financial indicators from raw stock data, which consists of Open, High, Low, Close, and Volume (OHLCV) values. These engineered features enhance the model's ability to detect trends, momentum shifts, and volatility patterns.

The implemented functions generate features in five key categories:

- **Trend Indicators:** SMA, EMA, MACD
- **Momentum Indicators:** RSI, Momentum
- **Volatility Indicators:** Bollinger Bands, ATR
- **Volume-Based Indicators:** Moving Average of Volume, OBV
- **Lagged Returns:** Past price values for time-series modeling

2.2 Trend Indicators

Simple Moving Average (SMA) The SMA is calculated as the mean of closing prices over a specified window:

$$\text{SMA}_t = \frac{1}{N} \sum_{i=t-N+1}^{t} \text{Close}_i \tag{1}$$

where N is the number of days in the window.

Exponential Moving Average (EMA) Unlike SMA, EMA assigns exponentially decreasing weights to older prices, making it more responsive to recent price changes:

$$\text{EMA}_t = \alpha \cdot \text{Close}_t + (1 - \alpha) \cdot \text{EMA}_{t-1} \tag{2}$$

where $\alpha = \frac{2}{N+1}$ is the smoothing factor.

Moving Average Convergence Divergence (MACD) MACD measures momentum by taking the difference between a 12-day EMA and a 26-day EMA:

$$\text{MACD}_t = \text{EMA}_{12}(\text{Close}_t) - \text{EMA}_{26}(\text{Close}_t) \tag{3}$$

MACD helps in identifying bullish or bearish momentum shifts.

2.3 Momentum Indicators

Relative Strength Index (RSI) RSI quantifies the magnitude of recent price changes to identify overbought or oversold conditions:

$$\text{RSI}_t = 100 - \left(\frac{100}{1 + \text{RS}}\right) \tag{4}$$

where RS is the ratio of average gains to average losses over a 14-day period.

Momentum Momentum measures the rate of change in price over a given period:

$$\text{Momentum}_t = \text{Close}_t - \text{Close}_{t-N} \tag{5}$$

This helps in detecting acceleration or deceleration in price movements.

2.4 Volatility Indicators

Bollinger Bands Bollinger Bands consist of an SMA with upper and lower bands derived from standard deviations of closing prices:

$$\text{Upper Band} = \text{SMA}_t + 2\sigma_t, \quad \text{Lower Band} = \text{SMA}_t - 2\sigma_t \tag{6}$$

where σ_t is the rolling standard deviation over a 20-day period.

Average True Range (ATR) ATR measures volatility by averaging the true range over a 14-day period:

$$\text{ATR}_t = \frac{1}{14} \sum_{i=t-13}^{t} \max(\text{High}_i - \text{Low}_i, |\text{High}_i - \text{Close}_{i-1}|, |\text{Low}_i - \text{Close}_{i-1}|) \tag{7}$$

ATR helps in assessing market volatility and setting stop-loss levels.

2.5 Volume-Based Indicators

On-Balance Volume (OBV) OBV tracks cumulative volume to indicate buying or selling pressure:

$$\text{OBV}_t = \text{OBV}_{t-1} + \begin{cases} \text{Volume}_t, & \text{if Close}_t > \text{Close}_{t-1} \\ -\text{Volume}_t, & \text{if Close}_t < \text{Close}_{t-1} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

A rising OBV suggests accumulation, while a declining OBV indicates distribution.

2.6 Lagged Returns

Lagged returns store past closing prices for time-series forecasting:

$$\text{Lag}_N = \text{Close}_{t-N} \tag{9}$$

This feature enables the model to learn from historical price patterns.

2.7 Integration with the Overall Project

Feature engineering is a fundamental step in the overall machine learning pipeline. These engineered indicators serve as inputs to classification models that predict future stock movements. By incorporating trend, momentum, volatility, volume, and historical return features, we enhance the predictive power of our trading strategy. These features will be used in the subsequent modeling and backtesting phases to assess profitability and risk management strategies.