Assignment 1 Report

- Kabyashree Dey(22B3013)

1. Normalization & Stationarity

I used a combination of Min-Max normalization and Z-score standardization, depending on the type of data.

- **Price Data (OHLCV)**: I applied **Min-Max scaling** to normalize prices and volumes between 0 and 1. This helps the DRL agent learn relative changes and patterns in price more effectively, especially since raw prices vary significantly across assets (e.g., AAPL vs GOOGL). Min-Max scaling also ensures that numerical magnitudes do not dominate the learning signal.
- **Technical Indicators**: For features like RSI, MACD, ATR, etc., I used **Z-score normalization** (StandardScaler), since these indicators often have different value ranges and are centered around zero or oscillate within specific bounds. Standardization helps in preserving the directional movement while making features comparable.
- **Stationarity**: I did not apply differencing or log transformation directly, but many of the engineered features (e.g., Price_change, momentum, ROC) are inherently stationary or close to stationary, as they focus on relative rather than absolute values. For a more rigorous model, differencing can be added before modeling if required.

I avoided over-normalizing to keep feature interpretability intact and to not distort indicators like RSI, which have meaningful thresholds (30/70 for oversold/overbought).

2. Chosen Technical Indicators & Rationale

The goal was to engineer features that capture **three key market dynamics**: **momentum**, **trend**, and **volatility**. I focused on intuitive, interpretable, and widely-used indicators in quantitative trading.

Trend Indicators

- **SMA (Simple Moving Average) & EMA (Exponential MA)**: These help detect underlying trends. SMA is smoother but slower to react, while EMA responds faster to price changes. They're useful for understanding bullish or bearish phases.
- **MACD (Moving Average Convergence Divergence)**: A classic trend-following momentum indicator that helps detect potential entry/exit points. It also includes a signal line and histogram for richer interpretation.

Momentum Indicators

- **RSI (Relative Strength Index)**: Measures the speed and change of price movements. Especially useful for identifying overbought and oversold conditions.
- **ROC (Rate of Change)**: Captures the percentage change in price over a fixed period, useful for identifying recent momentum spikes.

- **Stochastic Oscillator (STOCH_K, STOCH_D)**: Indicates momentum by comparing a security's closing price to its price range over a period.
- **Williams %R**: Similar to stochastic but inverted. Offers another way to catch overbought/oversold signals.
- **CCI (Commodity Channel Index)**: Measures price deviation from the statistical mean; good for identifying emerging trends or reversals.

Volatility Indicators

- **ATR (Average True Range)**: One of the most reliable measures of market volatility. Used in DRL for position sizing and risk adjustment.
- **Bollinger Bands** (Upper/Lower/Middle, Width, Position): Helps visualize volatility and possible breakout points. The width of the bands gives an intuitive volatility measure.

Volume-Based Features

- **Volume SMA**: Smooths out noise in raw volume data.
- **Volume Ratio**: Helps in identifying volume spikes which often precede major price moves.

Additional Price Features

- **Price_change**: Daily percentage return; acts as a stationarized input.
- **High_Low_ratio & Close_Open_ratio**: Simple ratios that capture daily intraday volatility and price sentiment.

Advanced DRL-Specific Features

To support the DRL agent's contextual learning:

- **Momentum (5/10 day)**: Multi-horizon short-term trend capture.
- **Volatility regime**: Ratio of short to long ATR; used to infer current market state (calm vs volatile).
- **Trend strength**: Absolute difference between MACD and its signal.
- **Mean reversion signal**: Captures how far price has moved from mean (SMA_20) in terms of ATR.
- **Risk-adjusted return**: Captures whether price movement is significant compared to recent volatility.
- **Relative volume**: Ratio of current volume to its 20-day mean.

3. Summary

This engineered dataset captures a rich variety of market signals across trend, momentum, and volatility domains. It avoids excessive complexity and is built using only pandas and numpy for transparency and reproducibility.

This structured feature space can serve as a strong foundation for reinforcement learning, enabling the agent to learn from a stable, interpretable, and diverse set of indicators. Features like RSI and MACD provide intuitive decision boundaries, while volatility-aware features like ATR and risk-adjusted returns ensure dynamic strategy adaptation.