Assignment 1 Report

Tharun Tej June 21, 2025

# Libraries Used

* yfinance: Used to fetch historical stock data from Yahoo Finance.
* pandas: For handling and transforming tabular data.
* numpy: For basic math operations and arrays.
* matplotlib.pyplot: To create graphs and plots.
* seaborn: For better looking visualizations.
* sklearn.preprocessing.MinMaxScaler: To normalize features between 0 and 1.
* pandas ta (or ta): For calculating technical indicators like RSI, MACD, etc.

# Normalization and Stationarity Methods

When we work with financial data like prices, the values can be very different for each stock or crypto. Some prices may be in the hundreds while others may be small numbers. To make them easier to work with and to help the model learn better, I used normalization.

I used Min-Max normalization because it is easy to understand. It changes all the values in a column to be between 0 and 1. This helps us compare different columns without one having more weight just because it has big numbers.

I also looked at Z-score normalization. This method tells us how far each value is from the average in terms of standard deviations. It is helpful when the data has outliers. Making the data stationary (i.e., removing trends) is important for time series models. I tried using differencing (value - previous value) on closing prices to remove trends. This helped make the data more stable and easy to model.

# Technical Indicators and Why I Picked Them

To understand the market better, I added some indicators. These are formulas that help us know if the market is going up, down, or is too risky. I used the Pandas and TA-Lib libraries for this.

Here are the 5 indicators I used:

* **RSI (Relative Strength Index)**: It helps us know if a stock is overbought (too high) or oversold (too low). It ranges between 0 and 100.
* **MACD (Moving Average Convergence Divergence)**: It tells us about the mo- mentum of the market and if a trend is changing.
* **Bollinger Bands**: It shows how much the price is moving around the average. If the bands are far apart, the market is more volatile.
* **SMA (Simple Moving Average)**: It is just the average of prices over a period of time. It smooths the curve.
* **EMA (Exponential Moving Average)**: Like SMA but gives more weight to recent prices. It reacts faster to changes.

I chose these because they are simple but useful. They cover momentum (RSI, MACD), trends (SMA, EMA), and volatility (Bollinger).

# Conclusion

After collecting more than 5 years of daily data using Yahoo Finance API for two stocks (for example: AAPL and TSLA), I applied these indicators and visualized how prices move with these features. I also normalized the data and split it into training, validation, and test sets without shuffling (to keep the time order). This data is now ready for use in training a RL model. I kept everything simple so that anyone can understand.