Dataset URL: API

Data is retrieved via Yahoo Finance API through yfinance library in Python, with S&P 500 constituents list of bases from: ttps://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies.

Description: This dataset contains daily stock data of every S&P 500 company for 5 years. The data captured for every company are Daily Price data (OHLC), Daily Volume, technical indicators (RSI, MACD, Bollinger Bands), fundamental metrics (P/E ratios, Market Cap), Quant factors (Beta, Market Returns), Volatility, Money Flow etc. By having exposure to all historical exchange data, we can analyse trends and price behaviour across both industries and market conditions.The dataset will contain inclusive daily trading data for all S&P 500 companies between January-2020 and December-2024 and it will include:

Dataset Overview:

Total Companies: 501

Total Observations: 622641

Date Range: 2020-02-02 to 2025-01-31

Number of Features: 76

Failed Tickers: 0

Data Points per Company (Avg): 1242.7964071856288

### Features Description:

* Price Indicators: ['Close: Daily closing price', 'Returns: Daily price returns', 'Log\_Returns: Natural logarithm of returns', 'Price\_Range: Daily high-low range', 'Price\_Range\_Pct: Price range as percentage of closing price']
* Moving Averages: ['MA\_X: Simple Moving Average (X=5,10,20,50,200 days)', 'EMA\_X: Exponential Moving Average (X=5,10,20,50,200 days)', 'Returns\_Xd: X-day price returns']
* Volatility Metrics: ['Volatility\_Xd: X-day rolling standard deviation of returns', 'Volume\_MA\_Xd: X-day volume moving average', 'BB\_Width\_X: Bollinger Band width for X-day period']
* Technical Indicators: ['RSI\_X: Relative Strength Index (X=9,14,25 days)', 'MACD: Moving Average Convergence Divergence', 'Signal\_Line: MACD signal line', 'MACD\_Histogram: MACD - Signal Line', 'Momentum\_14: 14-day momentum', 'ROC\_14: 14-day rate of change', 'MFI\_X: Money Flow Index (X=14,28 days)', 'Channel\_Width\_X: Price channel width (X=20,50 days)']
* Volume Indicators: ['OBV: On-Balance Volume', 'Volume\_Ratio: Volume relative to 20-day average', 'Volume\_StdDev: 20-day volume standard deviation']
* Fundamental Features: ['PE\_Ratio: Price to Earnings ratio', 'PB\_Ratio: Price to Book ratio', 'Dividend\_Yield: Annual dividend yield', 'Profit\_Margin: Company profit margin', 'Beta: Stock beta coefficient', 'Enterprise\_Value: Company enterprise value', 'Forward\_EPS: Forward earnings per share', 'Trailing\_EPS: Trailing earnings per share']
* Market Features: ['Market\_Return: S&P 500 daily returns', 'Market\_Volatility: S&P 500 20-day volatility', 'Rolling\_Beta: 60-day rolling beta coefficient', 'VIX: CBOE Volatility Index', 'VIX\_MA\_10: 10-day moving average of VIX']

Research Questions:

1. How effective are deep learning models (especially CNNs) in enhancing the precision, recall, and F1-score of technical pattern recognition in stock price movements compared to traditional rule-based approaches?
2. How does the prediction performance of multi-LSTM/RNN models that employ a combination of technical indicators compare to conventional moving average crossover strategies on a short-term (1-3 days) price movement forecasting task, measured by RMSE, MAE and directional accuracy?