Project Overview

The goal of this project is to analyse and predict the trading positions (buy, sell, hold) of Nifty50 companies using historical stock price data and technical indicators. The project involves several steps:

Data Extraction: Extract data for any Nifty50 company within a specified time frame with a ticker size of 1 day.

Technical Indicators: Calculate various technical indicators from scratch.

Feature Construction: Construct new features from the initial set of indicators using statistical techniques.

Trading Decision: Use the indicators to make trading decisions on a 1-day ticker.

Modeling: Fit a multivariate logistic regression model from scratch to predict trading positions.

Evaluation: Report various metrics for classification, including F1-Score, Accuracy, and AUC-ROC Score.

Data Extraction

For this project, historical stock price data for Nifty50 companies will be extracted from a financial data provider such as Yahoo Finance. The data will include the stock's open, high, low, close prices, and volume for each trading day within the specified time frame.

In my case I chose Reliance industries as my company to do analysis on as one of the largest companies in India by market capitalization, making it a prominent player in the Nifty50 index.Reliance operates in various sectors, including energy, petrochemicals, textiles, natural resources, retail, and telecommunications. This diversification can provide a comprehensive view of the market conditions.Apart from all these I personally did some fundamental analysis inorder to know if it is good for long term investments and I found the fundamentals to be very good for the company and it is a primary sector company which grows along with the GDP of the country an I found that Reliance has very little fluctuations in the market compared to other companies.

The technical Indicators I used are EMA,SMA,MACD,BollingerBands,OBV,StandardDeviation, and I read some blogs about how these indicators give signals and analysed them and for the final call I found EMA and SMA are very good indicators to go with.

If EMA>SMA it is a buy call and if EMA<SMA it is a sell call if EMA and SMA don't have a huge difference (less than 5) it is a hold call.

I also calculated Calls given by various indicators standalone to get an idea how these indicators work in real life if EMA ,SMA>CLosingPrice it is a buycall or else it is a sell.If MACDHistogram is +ve it is a buy call or it is asell if it close to zero it is hold. Similarly I did it with other indicators as well.

After calculating all the indicators I found that these have huge variations in their values which will give a very high bias while training machine learning modelso I used Mean Normalisation Technique to make them comparable.

For MachineLearning Model I used 5 indicators to train and I made 3 models for buy,hold,sell to predict which will give their respective probability.

I split the data into 3 types i.e according to Call and train them separately.

I calculated accuracy and F1 score and got 75% and 81% respectively.

EMA = (Close - EMA(previous day)) * (2 / (N + 1)) + EMA(previous day)
SMA = (Sum of closing prices for N periods) / N
MACD Line = 12-day EMA - 26-day EMA
Signal Line = 9-day EMA of the MACD Line
MACD Histogram = MACD Line - Signal Line
Middle Band = SMA
Upper Band = SMA + (Standard Deviation of price * K)
Lower Band = SMA - (Standard Deviation of price * K)
On-Balance Volume (OBV):
OBV = Previous OBV + Current Volume if Current Close > Previous Close
OBV = Previous OBV - Current Volume if Current Close < Previous Close
OBV = Previous OBV if Current Close = Previous Close

For cases where data was not available for the previous days I took Closing Price data for that in all above technical indicators

For MLModel I used Multivariable Logistic regression where variables are new features formed from technical indicators.

Loss=-1/n(ylogy+(1-y)log(1-y))
I used gradient descent to optimise parameters for backpropogation.