Low Level Design (LLD)

INVESTMENT PREDICTIONS

RITESH RAJURKAR

**Abstract**

Stock market plays a pivotal role in financial aspect of the nation's growth, but stock market is highly volatile and complex in nature. It is affected by significant political issues, analyst calls, news articles , company's future plans of expansions and growth and many more. Hence, any investor would be interested in understanding the stock market overtime and how the factors mentioned above affect the behavior of the stock market. On Every business day, millions of traders invest in stock market. Most of these investors lose money and others gain. However, considering any trading day, loss or gain is absolutely inconsistent. The demand to predict stock prices are extremely high, hence is the need for stock market analysis. This project is focused on analyzing a stock of any given company based on statistical technical indicators. Some of these indicators are deterministic in nature and the remaining are probabilistic. The objective of this project is to minimize the risk of loss in every trade thereby maximizing the profit.

# 

# Introduction

## Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Investment prediction system. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli.

**Introduction**

The world is flooded with information and we are only concerned about the information relevant to us. The information available can be classified based on various fields like engineering, arts, science, history, sports, geography, economics. In such situations it becomes important for us to boil down to the information relevant to us.

Especially when it comes to the field of finance, the information available is not well organized and they are just information of present state. The stock market can be viewed as a platform where almost all major economic transactions in the world occur at a dynamic rate called the stock value which is based on the market equilibrium. A correct prediction of this stock trend beforehand can earn huge profits.

In this project, two approaches are implemented to achieve the prediction. One of them is to use various technical indicators. These technical indicators are implemented to anticipate future changes in the stock trends. A technical indicator is based on mathematical calculations that can be applied to a stock's price, volume, or, even, to another technical indicator. The deterministic nature of the technical indicators will provide a strong evidence to foresee the stock market. The technical indicators implemented in this stock market prediction tool will back the predictability strongly. The second approach is a more probabilistic, which is based on the Hidden Markov Model.

This model is more suitable for dynamic systems and has been broadly used in pattern recognition problems. It is centered on the statistical methods and has the ability to handle new data robustly. In this model, based on the historical datasets that complement with present day's stock price behavior, the two dataset's are interpolated with suitable neighboring price elements. This results in a forecast for tomorrow's stock trend of the variable of interest. Stock market can be analyzed based on fundamental analysis which is about financial results of the particular company which we are analyzing and also important financial news about the company. This brings us down to news analysis in which we need to parse all the relevant documents and news about the company to extract required information based on which we can make a decision to invest Stock market runs on a well-known theory known as “Theory of Demand and Supply”. It states demand is always inversely proportional to supply. Hence, as demand increases buyers take over the stock market and market turns bullish. On the other hand, as soon as supply comes in sellers take over the market and stock prices start falling drastically. Stock market can also be analyzed based on “Head and Shoulder Theory” according to which, Highest price of any stock is considered the head and the next peak high is considered shoulder. Most of the indicators which we will see in this report are statistical in nature. Hence, there are a few assumptions which are made in the design phase of every technical indicator. For my analysis I have used data from Indian stock market which includes Bombay Stock Exchange and National Stock Exchange.

## Key Terms

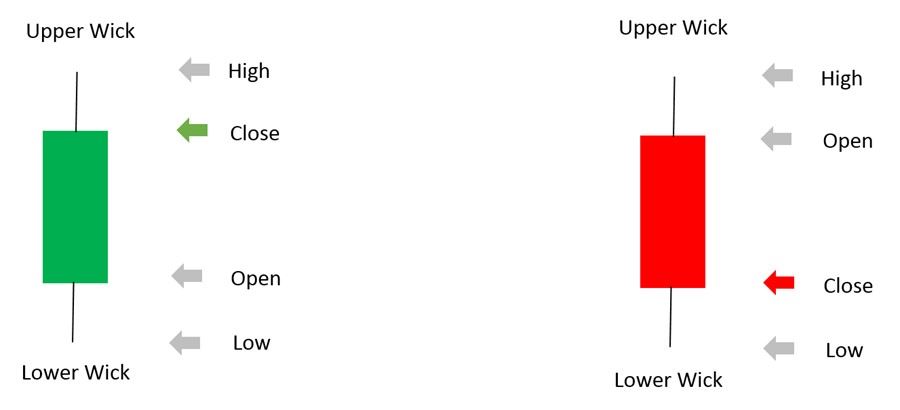
Interval Millions of trades take place every second. Hence, for analysis we need to classify these trades based on the interval at which they took place.

We can divide these trades into intra-day and long term intervals. Intra-day can be sub-classified into 1 minute, 2, 5, 10, 15, 30 and 60 minutes. Long term intervals can be classified into daily, weekly, monthly and so on.

Tick Prices At every interval any stock will have 4 types of prices associated with it. High price, Low price, Open Price and Close price.

High price is the highest value at which it was traded in that particular interval. Low price is the lowest price which it reached in that interval. Open price is the first trade which took place at that particular interval and Close price the price at which the stock was last traded at that interval. Trend At any instance of time, any stock will have either higher demand than supply or lower demand than available supply. Hence we can classify the Trend into two types namely Bearish and Bullish. A stock is said to be in Bullish Trend if it has higher demand than its supply at that instance of time. If the stock has higher available supply when compared with demand, it is said it be in Bearish Trend.

Data Representation Data need to be represented in some format before we can start analyzing it. In this section we will discuss how can we represent data for better interpretation.



Japanese Candle Figure represents two different candle sticks. The green candle represents gain, and red candle represents loss. This candle stick always contains all attributes associated with an interval (Example: tick prices). Since it represents an interval what we are looking at, we will refer to a candle stick as an interval in the rest of this report. Line Graph Line graph is a graphical way of representing data points in stock market based only on close price or current trading price of the stock. Line Graph Accuracy Accuracy is defined as the result of number of right decisions made divided by total number of decisions.

Accuracy based on Close Price When we calculated accuracy, if we relied on next interval's close price for decision verification, we have mentioned it as Accuracy based on Close price in the rest of the report. When we make a Buy call, to verify based on close price, we check if next interval's close price is above current interval's close price. If so, we consider the decision as a right decision. Also, if we make a Short sell, to verify the decision based on close price, we check if next interval's close price is lesser than that of current interval's close price. If so, we consider the decision as right decision.

Accuracy based on Low/High Price When calculating accuracy, if relied on next interval's low/high price for decision verification, we have mentioned it as Accuracy based on high/low price in the rest of the report. When we make a Buy call, to verify the decision based on high price, we check if current interval's close price is lesser than next interval's high price. If so, we consider the decision as a right decision. Similarly, if we make a Short sell, to verify the decision based on low price, we check if next interval's low price is lesser than current interval's close price. If so, we consider this decision as right decision. Short Sell Short sell is a type of trade where we sell the stock first, before buying it and buy it at a later point of time. Investors perform this trade if they think stock market is going to follow Bearish trend. Buy Call Buy call is a type of trade where investors invest by buying the stock first and selling it at a later point of time.

They perform this trade if they think, the stock in which they are investing will follow Bullish trend and can yield them profit. Technical Indicators Technical Indicators are the properties associated with any stock based on its tick prices. Calculation of each indicator is mentioned in the following subsections. Interpretation of each of the indicator is explained in detail in section Simple Moving Average (SMA) Simple Moving Average is calculated exclusively based on close price of the stock which we are trying to analyze.

For instance if we need to calculate SMA of 'x' intervals we need to get close prices of previous 20 intervals and divide it by 'x'. Hence the first available SMA value will correspond to xth interval. To calculate we add all close prices starting from the current interval looking back for n number of intervals. In this case n stands for number of intervals for which we need to calculate SMA.

EXPONENTIAL MOVING AVERAGE (EMA) Exponential Moving Average is also calculated based on close price of the stock which we are analyzing. If we need to calculate EMA of 'x' intervals, we first need to calculate SMA till x th interval and for every subsequent interval we calculate EMA based on the following formula.

EMA = Hence, the first available EMA value will be corresponding to xth interval. EMA moves hand in hand with price of stock. Higher the number of intervals we choose to calculate EMA, higher the stability of EMA. Hence, we chose 2, 5, 10, 20, 50 and 100 intervals for calculating EMA.

Relative Strength Index (RSI) Relative Strength Index is calculated based on SMA and close price of the stock for the given interval. We must get familiar with the following terms to better understand the calculation of RSI: Gain, Loss. If the close price of a stock at a given interval is greater than its open price, the stock resulted in Gain, vice versa it resulted in Loss. Here are the formulae to calculate RS and RSI.

RS = Average Gain/ Average Loss RSI indicates the strength of the current trend. Higher the value of interval we choose, we get stable RSI values. We need to find out a threshold value.

If RSI falls below its threshold, it is an indication of sellers taking over buyers. If RSI value rises over its threshold, it indicates that buyers are taking over sellers and stock prices will go high.

Bollinger Band (BB) Bollinger Band is calculated based on Standard Deviation and close price of the stock at a given interval. Bollinger Bands are calculated based on the following formulae:√ 𝑈𝑝𝑝 𝐵 [2 ] 2 [2 ] 𝐿 𝑤 𝐵 [2 ] 2 [2 ] Bollinger Band’s values provide an insight on how much more the stock can rise if it is in Bullish trend. Or, how much can it fall, if the stock is in Bearish trend. Fast Stochastic (FS) Fast Stochastic is the technical indicator which involves two values, namely, %k and %d. For a given interval these values are calculated based on current close price of the stock, lowest low price in the look-back period and highest high price of the stock in the look-back period. Lowest low and highest high are the lowest price and highest price at which the stock was traded in the given look-back period with respect to given interval respectively.Here are the formulae to calculate %k and %d: 𝐿 𝑤 𝐿 𝑤 𝐿 𝑤 𝐿 𝑤 At a given interval %k’s value falls below %d’s value, market takes Bearish trend. Otherwise, if %k’s value rises above %d’s value it indicates Bullish trend. However, it is very difficult to identify how long will the trend predicted from FS is valid. Moving Average Convergence Divergence (MACD)

This indicator is calculated based on exponential moving average. To calculate MACD we first need to compute EMA of 9 intervals and EMA of 26 intervals. Now starting from 27th interval till 35th interval subtract EMA (26) from EMA (9) and store the result for further processing.[4] This value represents upper MACD indicator. With these values now calculate EMA of previously stored results starting from 36th interval. This represents lower MACD indicator. Minimum number of intervals needed for MACD is 35 and is its only limitation. [5] Whenever lower MACD value falls below upper MACD value it shows a trend reversal from Bullish to Bearish. If lower MACD value rises above its upper MACD value it indicates a trend reversal from Bearish to Bullish.

Williams Average (WA) Williams Average is calculated based on current close, highest high and lowest low . Here is the formula to calculate Williams Average: In order to calculate this indicator we first subtract current close from highest high and store the result and call it HC. Now we subtract lowest low from highest high, store it and call it HL. With these results we now calculate Williams Average by the following formula. WA indicates oversold and overbought conditions. Whenever it reaches close to 0 it shows oversold condition. If WA value is closer to 100 it shows overbought condition. We need to identify the optimal threshold for which this indicator gives accurate result. Fibonacci Series Fibonacci Series is calculated using the following formula: √ Fibonacci series helps us to identify support level or resistance level of a given stock. Whenever we make a decision based on certain indicator we can use this indicator to identify we made the right decision before making a trade.

Rate Of Change (ROC) To calculate ROC we need to know the current trading price of the current interval and close price of the interval which is look-ahead parameter intervals previous to current interval. Here is the formula to calculate ROC. [4] Rate of change is directly proportional to the trend of the stock market.

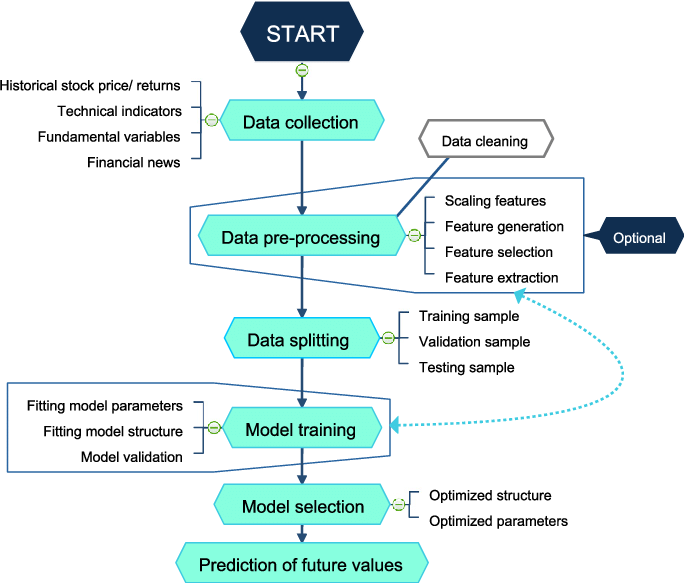
If ROC is lesser than 20 it is an indication that the market is in bearish trend. If ROC is greater Page | 13 than 80 it is a strong signal to buy stocks. Artificial Neural Networks (ANN) This model is a multi-level perceptron model which feeds forward the result of every computation to the next level of computation. The author has modified the approach of feed-forward model by adding pseudo input variables into ANN. First version of ANN propagates back and forth to provide final results. Author in his approach has argued that, by modifying the model he could minimize the back and forth movement of the data to increase optimization of computer’s computational power. ANN’s are always very less susceptible to noise points in the data. If a model is trained based on noise points accuracy of the prediction would decrease drastically due to increase in false positives. Using this model author is making the decision to hold, sell or buy stocks at given time. He used a Genetic algorithm to select the data from the given data set.[20] This selected data will be given as an input to ANN which would classify the labels based on the weight of computed results. In his results he proves that 2 layer ANN is more feasible than a single layer perceptron. The author [9] has improved the model by embedding pattern matching with basic ANN. Pattern matching algorithms are always highly susceptible to noise.[13] However, for given patterns the algorithm will detect the re-occurrence if any. Hence, by embedding this advantage with ANN author has opened the doors of fortunes. This improved model can now be used for prediction on data that contains extreme noise (stock market often is the case).

He has also proved the same with his results in the publication. Hidden Markov Model (HMM) Hidden Markov Models are statistical models that are used to determine the most likely sequence of occurrence given a sequence of data. This model consists of transition matrix, observation matrix and initial set of observation distribution, denoted as A, B and pi respectively.

This model is used to uncover hidden information behind any set of sequences that has happened. In stock market ticks of respective intervals are taken into consideration for computing initial set of observations and observation distribution. Based on the computed results we then calculate the transition matrix. It is the transition matrix, which author relied on to calculate the likeliness of the trading sequence. When he trained this model with ticks whose results we already know since they are training records. He then calculated its likeliness. Now he calculates likeliness for a sequence whose results we don’t know yet. It is based on this likeliness; we predict the future trend for that particular stock.

The author has used tick prices of current interval as 4 of his observations. He also considered close price of next interval, relative to the current interval, as fifth observation for training his model. Based on these data he computes A, B and pi as discussed previously. This model now can be used to predict next most likely observation.

# Model training/validation workflow



# User I/O workflow

