CHAPTER I

INTRODUCTION

1.1 Overview

This thesis deals with the development of predictive data mining techniques which help the investors to discover hidden patterns from the historic financial time series data to forecast the stock market. The major contributions of this thesis are to maximize the prediction capability of financial stock market using time series quantitative technical analysis, particularly by formulating variations of existing technical indicators; to devise combinational algorithms by combining various technical indicators and the variations of the indicators formulated; to devise a unique metric based on "Correct Profitable Trade Signal - CPTS" for evaluation of the stock market predicting algorithms; to study the performance of the Moving Average Crossover (MAC) Algorithm [16], [86] on the historic time series data in the perspective of percentage of CPTS produced and the percentage of annual Returns on Investment (ROI) generated, which is taken as a benchmark to compare the performance of MAC with the proposed combinational algorithms.

This introductory chapter discusses the research issues in predicting the stock market. In financial practice it is not the question whether it is possible to forecast, but how the future path of a financial time series can be forecasted. In academia, however, it is merely the question of whether series of speculative prices can be forecast rather than the question of how to forecast. Therefore practicing financial professionals and academicians have proceeded along different paths in studying financial time series data.
Quantitative analysis is the process of determining the value of a security by examining its numerical, measurable characteristics such as revenues, earnings, margins, and market share from the time series. The fundamental [10] and technical analyses [62] are the techniques developed in financial practice to forecast financial time series. The purpose of fundamental securities analysis is to find and explore all economic variables that influence the future earnings of a financial asset. The Fundamental Analysis is based on macroeconomic data and the basic financial status of companies like money supply, interest rate, inflationary rates, dividend yields, earnings yield, cash flow yield, book to market ratio, price-earnings ratio and lagged returns.

The Technical Analysis is based on the rationale that history will repeat itself and the correlation between price and volume reveals market behavior [86], [78]. Technical analysis is the study of past price movements with the goal to predict future price movements from the past. Both these analyses are intended to give advice on what and when to buy or sell. These techniques have the probable capability to help the investors in the investment decisions by looking into the new hidden patterns and opportunities.

Prediction is made by exploiting implications hidden in past trading activities and by analyzing patterns, which are essentially based on Technical Analysis, as most of the parameters of Fundamental Analysis are either static and well known or global in nature implying that they contribute only a little in predicting the behavior of individual stocks.

As long as financial markets have existed, people have tried to forecast them in the hope that good forecasts would bring them great fortunes. Technical analysis has been a popular and heavily used technique for decades already in financial practice. When
multiple non-correlated time series technical analytics are combined, there is a possibility to improve the predictive power [84]. The combined performances of selective financial indicators are studied and the combinational patterns that maximize profitable trades as well as result in high return on investment are explored.

Competitive business pressures and a desire to leverage existing information technology investments have led many financial investment firms to explore the benefits of data mining technology. As of now, more than 70 per cent of trades in the US stock market are done by automated computer programmes using data mining and predictive technologies.

Many fund management firms have invested heavily in information technology to help them manage their financial portfolios. Over the last three decades, large amounts of historical data have been stored electronically and this volume is expected to continue to grow considerably in the future. Yet despite this wealth of data, many fund managers have been unable to fully capitalize on their value. This is because information that is implicit in the data for the purpose of investment is not easy to discern.

For example, a fund manager may keep detailed information about each stock and its historic data but still it is difficult to pinpoint the subtle buying patterns until systematic explorative studies are conducted. This issue is addressed by applying data mining technology using quantitative analytical techniques to help to discover previously undetected patterns present in the historic data to determine the buying and selling points of equities.
Time series quantitative technical analytics are well documented and are used to perform technical analysis on stocks [83]. However, a clear cut strategy is lacking to generate buy and sell conditions on individual stocks. As these analytics are done upon the same data, there is a possibility of having high correlation between multiple analytics [29]. This is the basic motivation to study combinations of uncorrelated analytics in order to improve the predictive power.

The effectiveness of any financial time series forecasting algorithm cannot be measured simply by means of the number of trade signal generated which reflects the stock market trend correctly, because all the correct trade signals do not yield net profit which results in making good Returns On the Investment (ROI) of money.

As there is no minimum guarantee for all the trade signals generated by a stock market predicting algorithm to yield profit, an innovative performance evaluation metric, CPTS, is evolved in this thesis. The generation of a complete set of one buy signal, one sell signal and one neutral signal is considered as closed trade signal resulting in either profit or loss. The CPTS is devised based on this closed trade concept and provides a comprehensive meaning to the trade signals generated.

This thesis investigates the performance of combinatorial quantitative algorithms to find out the best combination patterns that have the predictive power of high probability of profitable trades and high percentage Returns on Investment (ROI) of money.
The accomplishments of this work include:

- Discussing the features of stock, stock market, financial time series, and predictive data mining.
- Discussing the features of Efficient Market Hypothesis (EMH), its critics, effects of long-term memory component, arbitration, fundamental and technical analysis.
- Discussing various existing combinational quantitative technical time series algorithms to predict the stock market.
- Designing and devising new variations of existing financial indicators.
- Designing and developing novel combinational quantitative technical analytics to increase the percentage of profitable buy / sell signals produced and the percentage of annual returns generated.
- Devising an innovative metric “Correct Profitable Trade Signal - CPTS” for evaluation of the performance of combinational quantitative technical algorithms.
- Performing subjective back testing of the new combinational algorithms and comparing them against the performance of the Moving Average Crossover Algorithm.

1.2 Organization of the Thesis

This thesis begins with a review of the research area which makes a prologue on significance of Data Mining, stock, stock market and financial time series data in the
chapter II followed by details of the financial practices, memory in the stock returns, the
types of informational efficiency, the arbitrage opportunities arising from the difference
between fundamental and market values of the stock prices, the efficient markets
hypothesis (EMH) [9] and its critics in chapter III. Chapter IV discusses the literature
survey of the combinations of uncorrelated existing analytics to predict successfully the
price movements of stocks. The limitations encountered in the behavior of these
techniques are identified that motivated the formulation and development of new
combinational algorithms recorded in chapter V of this work. The Moving Average
Crossover (MAC) [86] is a simple straightforward smoothing algorithm which is taken
as a benchmark to compare the percentage of profitable buy / sell signals generated by
the proposed combinational algorithms is discussed in this chapter. Also the variations of
the existing indicators designed to improve the predictive power of the combinational
analytics is presented in this chapter V. The details of the combinational time series
quantitative algorithms identified in this thesis are given below in detail.

**Technique 1: Time Series Quantitative Algorithm 3 – TSQA3**

TSQA3 analyzes the stock price using the optimized trading rule, which is
evolved by combining the, Chaikin Money Flow Indicator (a confirming indicator) with
the Relative Strength indicator (RSI), Stochastic Momentum Index, Bollinger Bands
Calculation and RSI Signal. Stephen C.W. Chu [90] combined broad range of technical
indicators, including Moving Averages, RSI, Stochastic and Bollinger Band. The overall
prediction capability of this algorithm is enhanced by combining the Chaikin Money
Flow indicator. The Chaikin Money Flow indicator is calculated from the daily readings
of the Accumulation / Distribution Line [90]. The basic premise behind the Accumulation
Distribution Line is that the degree of buying or selling pressure can be determined by the location of the Close relative to the High and Low for the corresponding period (Closing Location Value) [47].

**Technique 2: Time Series Quantitative Algorithm 4 – TSQA4**

TSQA4 is formed with a variation of this oscillator called Stochastic Momentum index combined with the confirming indicator, Chaikin Money flow Indicator, Bollinger Bands Calculation, Relative Strength indicator and Stochastic Momentum Index Signal for achieving better results. Piotr Lipinski [76] proposed a new trading rule by combining Stochastic Oscillator with Relative Strength Indicator. This work was not focused on profitable buy / sell signals generation, but proposed a technique of generating warning signals for specific situations relate to the rare circumstances of exceptional raises or drops of share prices. The Stochastic oscillator, a popular and dynamic indicator is based on the premise that during an upward trading market, prices tend to close near their high and during a downward trading market, prices tend to close near their low.

**Technique 3: Time Series Quantitative Algorithm 5 – TSQA5**

TSQA5 is formed by combining True Strength Index (TSI) with other financial indicators mentioned below to improve the percentage of profitable signals generated. The TSI is a variation of the Relative Strength Indicator. The TSI uses a double smoothed exponential moving average of price momentum to eliminate choppy price changes and spot trend changes. The TSI combines the leading characteristics of momentum calculation with the lagging characteristic of a moving average to create an indicator that reflects price direction and is in synchronization with market turns.
True Strength Index Signal is combined with Stochastic Momentum Index, Bollinger Bands Calculation, Relative Strength indicator, and Chaikin Money Flow indicator yielding TSQA5.

Chapter VI provides the back testing strategy and the experimental results obtained by applying the combinational logic. An innovative metric for evaluating combinational quantitative technical analytics, CPTS is presented in this chapter. The results of the percentage of CPTS generated and the ROI obtained by the combinational algorithms are compared amongst themselves and with Moving Average Crossover algorithm which is chosen as a bench mark.

Chapter VII concludes the thesis and suggests way for the future course of the work as well. The main contribution of this work is devising meaningful and reliable models of financial time series from a collection of popular technical indicators by means of evolutionary combinational algorithms. This analysis explored the technical ways of predicting financial markets using quantitative analysis. Experimental results confirm that the proposed combinational quantitative analytics outperforms over moving average crossover algorithm.