Chapter 2

Survey of Literature

2.1 Introduction

Increase in economic globalization and evolution of information technology, has generated and accumulated an unprecedented volume of financial data. This data is mainly used to keep track of company’s business performance, monitor market changes, identify customers increased needs, and support financial decision making. The rapidly growing volume of data has far exceeded the ability of analyzing manually. Hence there was a need for automated approach to analyze the data effectively and efficiently. With the advent of database technologies, handling massive data has become a reality. Moreover, Data Mining techniques have been designed to uncover hidden patterns and predict future trends and behaviors in financial markets. Before we go further in to our research work we introduce the essential elements and concepts of Stock Markets.

2.2 Background on Stock Market

Over a time period Stock Markets have become the very symbol of commerce in the modern world for a country. They are truly unique in their scope, and in the complexity of the number of transactions they handle each day. The economy of the world relies on the stock exchanges to facilitate even trade in the stocks of companies.

2.2.1. Stock Market Definitions

Market Value:

This applies to the size of a company. Market value is also referred to as market capitalization or market cap for short. To figure out a company’s market cap, you must multiply the outstanding shares of the companies’ times the current
share price. The market cap of a company depends on the business it does in the market for a specific period. The categories are: 1) Micro-Cap: companies with market caps under $250 Million. 2) Small-Cap: companies with market caps between $250 Million and $1 Billion. 3) Mid-Cap: companies with market caps between $1 Billion and $5 Billion. 4) Large-Cap: companies with market caps greater than $5 Billion.

Book Value:

Book Value Take current total assets of a company minus total liabilities. Then divide the result by the current number of outstanding shares. This will give you the book value per share of a company. A company should never sell for less than this amount, and if it does, you may have found a bargain.

2.2.2. Stock Market Indicators

A stock market indicator refers to anything used to project future financial or economic trends. There are a number of factors at play when it comes to the rise and fall of stock prices on any given trading day. Investors who are able to grasp what these factors are will be more likely to achieve success when buying and selling stocks in either a bull market or a bear market.

Stock market indicators are extremely important in the process of determining the viability of purchasing stocks of a particular company. Investor should be able to understand, what the indicators are saying about a particular firm's stock value or growth prospects. Only then he will be able to make solid decisions, which are sure to bring in profits over the long term.

What follows is a list of some of the widely used economic indicators as well as definitions for each:

1. Market Cap:

   Refers to the total dollar value of all outstanding shares.
2. Price/Earnings (P/E) Ratio:

It refers to valuation of a firm's current share price compared to its per-share earnings. Typically an earning per share (EPS) is calculated by using the previous four quarters. A high P/E usually symbolizes significant projected earnings in the future.

3. Return on Equity (ROE):

This refers to the measure of a company's profitability, particularly as compared to other firms within the same sector. One way to determine ROE involves subtracting preferred dividends from net income and then subtracting preferred equity from shareholders equity.

4. Dividend Yield:

Refers to the income produced by a share of stock. The annual dividends per share divided by the price per share produces the dividend yield.

5. Price-To-Book Ratio:

Refers to the process of comparing a stock's market value to its book value. It is determined by dividing the current closing price of the stock by the latest quarter's book value. If there is a lower price-to-book ratio, one of two things could be the case. First, it could mean that the stock is undervalued. Or, second, it could instead mean that the company is not in the best of health.

2.2.3. Stock Market Strategy

When it comes to investing in the stock market, investors should know when to hold on to stocks and when to unload them.
Most financial experts believe that the buy-and-hold strategy, which requires investors to buy stocks and then keep them for the long term is the best method for ultimately making money on the stock markets. The rational behind this strategy is that, while the markets will likely experience ups and downs stemming from numerous factors, over time the stock markets tend to push upwards. This means that those who use the buy-and-hold strategy stand to make money over time. While there are many experts who still hold to this strategy, others point to some of the more catastrophic stock market crashes of the past as proof that investors can literally lose everything they had gained in a bull market to the impact of the bear market.

Experts say that bear markets materialize about every four or five years and bear markets can easily eliminate gains made in bull markets. History has proven this over and over. Rather than adopt a buy-and-hold strategy, some financial professionals recommend that investors take a more sophisticated approach to buying and selling stocks. This necessitates monitoring market conditions and making changes as fluctuations in the markets warrant change. What it doesn’t mean is making change just for the sake of making change.

Investors should also be wary of the advice they adhere to. While experts do provide a wealth of information on the markets and on individual stocks, it is up to investors to sift through the data and ultimately make the decisions. Some investors choose to go with a broker so as to bypass the pressures of managing their own stock portfolios. Doing so requires them to look around for a good broker, one who has a proven methodology and a solid track record.

2.2.4. Stock Market Trends

Most people believe the stock market is a mystery wrapped up in more intrigue than a best-seller whodunit. And to the masses, the ups and downs of the markets are beyond comprehension. But these and other similar views are
generally based on misconceptions of what the stock market is, how it works, what factors impact it and who participates in it.

From experience we know that investors may temporarily pull financial prices away from their trend level. Over-reactions can occur so that excessive euphoria drives prices unduly high, just as undue pessimism can push them down too far. New theoretical and empirical arguments have been put forward against the notion that financial markets are efficient.

According to the efficient market hypothesis [32], it is only changes in fundamental factors, such as profits or dividends that ought to affect share prices. Something that no doubt helped to question the explanatory power of the efficient-market hypothesis was the stock market crash in 1987, when the Dow Jones index plummeted 22.6 per cent - the largest-ever one-day fall in the United States. This dramatic event demonstrated that share prices could fall even though nothing more fundamental appears to have happened; a thorough search failed to detect any specific or unexpected development that might account for the crash. It also seems to be the case more generally that many price movements are not occasioned by new information; a study of the fifty largest one-day share price movements in the United States in the post-war period confirms this. (2) Moreover, a number of studies have shown that price movements have occurred solely because the company in question has been included in or excluded from Standard & Poor's 500 index, without any new information about fundamentals.

There are some simple principles to keep in mind when trying to grasp the rise and fall of the markets, what some people refer to as stock market trends. First of all, people tend to invest in companies that reliably earn a lot of profits. So people are obviously more likely to pay top dollar for shares in a successful company then they are to pay for shares in a less successful company. But that is only part of the whole equation.
Interest rates also impact the value of stocks. On the one hand, if interest rates are relatively high, then stock prices tend to drop, because people are more likely to avoid the roller coaster ride that is the stock market and play it safe by purchasing bonds or leaving their money in the bank. On the other hand, if interest rates are relatively low, then stock prices tend to rise, because people are more likely to endure the roller coaster ride that is the stock market rather than play it too safe and earn little or nothing with their funds locked away in the bank.

Furthermore, a solid economy is generally good news for stocks, as there will be more money available to purchase stocks. But when the economy is sluggish, the markets, too, could slump as result.

Positive and negative publicity is another factor to consider. If a pharmaceutical company announces that it has just gained approval to bring its products to the market, then it’s safe to say that its stock prices will go up. But if the same company announces that its new product has not been cleared for commercial distribution, then the markets could react negatively to the news.

Stocks can also go up and down depending on geo-political happenings. For instance, the war on terrorism has at times had a negative impact on stocks, as nervous investors panicked.

Apart from the factors mentioned above there might be several others, which affect the decision taken by an investor.

2.2.5. Stock Market Timing

Millions of dollars can be made or lost on the stock markets, which makes the subject of stock market timing such a compelling one for brokers, economists and casual investors. But with so many software, books and websites currently offering plenty of conflicting advice, how is an investor to know which strategy to
employ? The answer, quite simply, is research. Stock market timing is a science, albeit an inexact one. Different experts employ different metrics and focus on certain indicators more than on others. So investors simply have to do their own research to determine which strategy best fits their needs.

Put simply stock market timing refers to using calculations to determine the best times to buy, sell or hold. Some solutions available can be used straight out of the box to do just that. Some out-of-the-box solutions offer non-professional investors the chance to play the stock markets relatively independently. Through these solutions, investors can select their stocks and establish how much they are willing to spend. These solutions can also monitor stock price variations, among other things, and inform them when and how many stocks to acquire and unload at given periods.

2.2.6. Stock Market Trading

There are two ways for the investors to trade in the stock markets. “Primary Market” is a part of the financial market, where enterprises issue their new shares and bonds to raise capital for their investments. “Secondary Market” is the one where financial instruments such as stocks, bonds are traded after they are initially offered in the primary market. Most of the trading is done in the secondary market. BSE, NSE, etc., facilitate this trade in an organized way.

The third market in finance, refer to the trading of stock exchange listed securities in the over-the-counter (OTC) market. These traders allow institutional investors to trade blocks of securities directly, rather than through an exchange. Many of the financial brokerage firms facilitate this trading.

An advanced trading strategy is “short selling”. Most of the investors who are involved in daily transaction of stocks are involved in this trading. It involves selling of stocks, that the seller does not own, or any sale that is completed by the
delivery of stocks borrowed by the seller. Short sellers assume that they will be able to buy the stock at a lower price than the price at which they sell.

2.3 Stock Market Movements

Though the Stock Market prediction task divides researchers and academicians into two groups those who believe that the market is predictable and those who believe that the market is efficient and corrects itself. There are several classical prediction methods in the history of financial markets. Some of them are:

2.3.1 Classical Theories of Stock Market Predictions

Efficient Market Hypothesis is an investment theory which states that ‘it is impossible to beat the market as all the relevant information is fully and immediately reflected in a security's market price, thereby assuming that an investor will obtain an equilibrium rate of return’. In other words, an investor should not expect to earn an abnormal return (above the market return) through either technical analysis or fundamental analysis. According to the EMH, stocks always trade at their fair value on stock exchanges, making it impossible for investors to either purchase undervalued stocks or sell stocks for inflated prices. As such, it should be impossible to outperform the overall market through expert stock selection or market timing, and that the only way an investor can possibly obtain higher returns is by purchasing riskier investments. EMH has emerged as a prominent theory in mid 1960s. Fama’s contribution in EMH is significant; his theory breaks it into three forms: Weak, Semi-Strong, and Strong EMH [8, 28]. In Weak EMH, only past price and historical information is embedded in the current price. This kind of EMH rules out any form of predictions based on the price data only, since the prices follow a random walk in which successive changes have zero correlation. The Semi-Strong form goes a step further by incorporating all historical and currently public information into the price. This includes additional trading information such as volume data, and fundamental
data such as profit and sales forecast. The Strong form includes historical, public and private information, such as insider information, in the share price. The weak and semi-strong form of EMH has been fairly supported in a number of research studies [75, 76]. However the recent research has shown that Efficient Market Hypothesis is far from correct.

Another classical theory of stock market analysis is “Random Walk Theory”. Application of it to finance and stock markets suggests that stock prices change randomly, making it impossible to predict stock prices. The random walk theory corresponds to the belief that markets are efficient, and that it is not possible to beat or predict the market because stock prices reflect all available information and the occurrence of new information is seemingly random as well. The random walk theory is in direct contrast to technical analysis, which contends that a stock’s future price can be forecasted based on historical information through observing chart patterns and technical indicators. Academicians cannot conclusively agree or contradict this theory as there are published studies that support both sides of the issue [2, 18, 23].

Several researchers in the field of economics and management sciences still are working on methods using fundamental analysis. Their main area of interest in stock markets is to identify the accounting attributes which can be used for fundamental analysis. Identifying historical accounting signals that can be used to improve the entire distribution of future returns earned by an investor, whether the investor is doing business in primary, secondary markets or is a short-seller [13, 1]. There were researchers who were with the view that the stock market does not reflect the current earning of the company fully and hence the future trends can not be predicted. They argued that the financial markets depend on too many factors which makes it impossible to be predicted by the current accounting attributes of a company [4, 9].
2.3.2 Approaches to Stock Market Prediction

The importance of the Stock Market for the economy, and the complexity of it in the business environment had led to analyzing the markets for better prediction of the trends. There are several different types of methods applied to predict the Stock Market returns. All these prediction methods can be grouped into two major categories:

1. Fundamental Analysis Methods.
2. Technical Analysis Methods.

2.3.2.1. Fundamental Analysis

These methods use fundamental data such as growth, the dividend payout of the company, the interest rates prevailing, the risk of investment, the sales of the company, the tax rates imposed by the Govt. etc. The methods are aimed at computing the real value of the asset of company on which the investor is trying to invest. Once these methods evaluate the value of the asset, it comes simple to identify whether to invest or not. If the value calculated is higher than the value it holds in the market then invest in it, otherwise avoid the investment. Since these methods use mostly values provided by the company the significance of the company, its ethical value in the market, and its sincerity towards its investor all these play an important role. Though these analysis methods are objective and logical, and less dependent on psychological factors they are not very popular. This is simply because getting authentic values of fundamental attributes of a company are quite difficult, and hence not much dependable. Usually investors use these methods for long-term investment plans.

People who use fundamental analysis argue that the only way to predict the future performance of a company is to carefully examine its current financial state of operations through detailed analysis of its financial statements. The goal of this analysis combined with the industry outlook is to determine a fair market value for the security. If the stock trades below this value, then it should be
bought, if it trades above this value, then it should be sold. Fundamental analysts choose to neglect short term fluctuations in favor of long term appreciation.

2.3.2.2. Technical Analysis

Technical Analysis Methods use technical attributes of the market for prediction. The technical data comprises of price, volume, highest and lowest price for a trading period, etc. The methods mainly uses trends in these attribute values to predict the market movements. Charts are drawn to identify often noticeable patterns. Studying these charts gives an idea of trading rules using these technical attributes. Technical analysts believe that the market is mostly controlled by psychological factors rather than any logic. Hence the careful study of the patterns of market movement sheds light on to what the other investors are doing, which in turn gives an idea of future movement of markets. Because of its nature of prediction these methods are highly subjective, even then these methods are very popular as the market is driven by human psychology, which is difficult to be controlled.

DOW Theory is more than 100 years old, believed to be the foundation for technical analysis. It is a theory proposed by several contributors, to the work of Charles H Dow published as a series of Wall Street Journal editorials during 1900 to 1902. Some of the most important contributors and followers of Dow Theory were William P Hamilton, Robert Rhea, E George Schaefer, and Richard Russell. Dow Theory has showed that there are, simultaneously three movements in progress in stock market. The major is the primary movement bull or bear market. It is shown that this primary movement tends to run over a period of at least a year. In the course of it there will be Dow’s secondary movement, represented by sharp rallies in a primary bear market and sharp reactions in a primary bull market. Concurrently with the primary and secondary movement of the market, and constant throughout, there obviously was, the underlying fluctuations from day to day (minor trends). Dow’s theory in practice develops
many implications. One of the best tested of them is that the two averages (known as Dow’s averages of gain or loss) corroborate with each other, and the scrutiny of the averages show that there are periods where the fluctuations for a number of weeks are within a narrow range, this is technically called “making a line”. When the two averages rise above the high point of the line, the indication is strongly bullish and the secondary rally is in a bear market. If the two averages break through the lower level, it is obvious that the market has reached the saturation point and a secondary bear movement in bull market follows. For many years the theory has successfully predicted the movement of the stock prices [33, 26, 7, 27]. With the use of technology and the research in the field of financial markets, different methodologies have evolved over the period.

Candlestick Charts are believed to be developed in 18th century by Japanese trader Homma Munchisa. They are similar to bar charts with the combination of line-charts representing the range of price movements over a given time interval. It is a simple representation mechanism of complex data on a single chart. Candlesticks are usually composed of the body, and an upper and lower shadow (wick). The wick illustrates the highest and lowest trade prices of a stock during the time interval represented. The body represents the opening and closing trade. These charts are used to represent more complex patterns conveying more information than the other charts. Study of these charts reveals, support level, resistance levels of a stock. They also specify detailed description of bullish and bearish reversal patterns. Easier understanding of these charts, makes it more useful even today [32].

Elliott Wave Principle is one of technical analysis methods in the domain of stock market prediction. Investors use this mechanism to forecast trends in the financial markets by identifying extremes in investor’s psychology, high and lows in price movement, and other collective activities [24]. In this current work we have used this method along with other methods.
These methods use historical data to analyze and find the patterns which are generally believed to be cyclic. The extraction of such knowledge from the patterns is applied on to the current data to predict the market trends.

2.4 Data Mining: an overview

With the progress in database technologies, various kinds of advanced data and information systems have emerged and are undergoing development to address the requirements of new application. These applications include handling spatial data (maps), engineering design data (design of buildings, system components, integrated circuits), hypertext and multimedia data (including text, image, video, and audio data), stream data (video surveillance and sensor data, where data flow in and out like streams), and World Wide Web (widely distributed information repository made available by the Internet). These applications require efficient data structures and scalable methods for handling complex object structures. In response to these needs, advanced database systems and specific application-oriented database systems have been developed. These include object-relational database systems, temporal and time-series database systems, spatial and spatiotemporal database systems, text and multimedia database systems, heterogeneous and legacy database systems, data stream management systems, and web-based global information systems [39].

2.4.1 Models and Patterns

As the type of applications differs, the type of structures the data mining algorithms handle for the data also differs, and as the input data has different databases depending on the applications, the output from these algorithms are mainly categorized into models and patterns. These models and patterns are structures that are estimated, matched and in general utilized for various data mining objectives.

Model is a global, high-level and often abstract representation of data. They are specified by a collection of parameters, which can be estimated from the
given data. Modes are further classified based on whether they are predictive or descriptive. Predictive models are used in forecasting and classification applications; whereas descriptive models are useful for data summarization. Autoregression, Markov models are some popular classes of predictive models and spectrograms, clustering are good examples for descriptive modeling techniques.

Pattern is a local structure that makes a specific statement about a few variables or data points. Spikes are patterns in a real-valued time series that may be of some interest, similarly regular expressions in symbolic sequences is a well-defined patterns. Matching and discovery of such patterns are very useful in many applications. As the patterns have easy and readily interpretable structure, they play a dominant role in data mining. Though a distinction between modes and patterns is defined, the interdisciplinary nature of data mining field has narrowed these differences. However there are model-based methods that are used to better interpret patterns discovered in data, thereby enhancing the utility of both structures in temporal data mining.

2.4.2 Mining techniques

Data Mining applies data analysis and knowledge discovery techniques under acceptable computational efficiency limitations, and produces a particular enumeration of patterns over the data [93]. Different type of application needs different type of data mining techniques to unearth the knowledge from the data that is needed for it. These techniques used are classified into the following categories based on the knowledge mined [39].

I. Association Rule Mining

It uncovers interesting correlation patterns among a large set of data items by showing attribute-value conditions that occur together frequently. Market Basket analysis is one such type of example.
II. Classification and Prediction

It is the process of identifying a set of common features and models that describe and distinguish data classes or concepts. The models are used to predict the class of objects whose class label is unknown. A large number of classification models have been developed under this category for predicting future trends of the financial markets. In section we will look at these models in detail.

III. Clustering Analysis

It segments a large set of data into subsets or clusters. Each cluster is a collection of data objects that are similar to one another within the same cluster but dissimilar to objects in other clusters. These techniques are used to identify stable dependencies for risk management and investment management.

IV. Sequential Pattern and Time-series mining

Models in this category look for patterns where one event leads to another latter event. These models are extensively used as many of the emerging real time applications, such as sensor networks, networking flow analysis, e-business and stock markets change their course due to events that occur at a particular time.

Apart from the conventional data mining techniques there are several advanced techniques using machine learning, fuzzy logic and several artificial intelligence, that are being used in modeling the above. As our research objective is to predict the stock market prices we are concentrating on different techniques used in classification and prediction methods.

2.5 Temporal data mining

As the stock market data is continuous time stamped data and is a time series data. For the purpose of prediction we need to use a huge data set. Temporal data mining is of latest origin concerned with data mining of large sequential data. It is useful in discovering qualitative and quantitative temporal
patterns in a temporal database or in a discrete valued time series dataset. Although there is no notion of time as such, the ordering among the records is very important and is central to the data description or modeling. Temporal data mining however is somewhat different with constraints and objectives than the traditional time series data. One main difference lies in the size of data sets and the way it is collected with little or no control over gathering process. Often the methods must be capable of analyzing large data sets. The second major difference lies in the kind of information that we want to estimate or unearth from the data like, trends and patterns in the data which are easily interpretable.

However, the objectives of temporal data mining are grouped in to five categories of tasks. They are:

i. Prediction

This model uses past samples of data and predicts the future values. In order to do this, one needs to build a predictive model for the data. If the data is static, auto regressive family of models are used to predict a future value as a linear combination of earlier sample values \([3, 5, 12]\). Linear stationary models like ARIMA models have also been found useful in many economic and industrial applications where some suitable variant of the process can be assumed to be stationary. There are many nonlinear models for time series predictions line neural networks \([30, 31, 11, 16]\).

ii. Classification

In this category, the sequences presented are grouped into predefined finite classes and the goal is to automatically determine the corresponding category for the given input sequence. There are several applications of this sequence classifications line in; speech recognition system to transcribe speech signals into their corresponding textual representation \([80, 81, 82]\). There are some pattern recognition applications in which, images are viewed as sequences. Recognizing handwritten words is an interesting application of this nature \([83]\).
Signature verification applications [84] also fall into this category. As in any standard pattern recognition framework all these applications use feature extraction step that precedes the classification step. Sequence classification applications have seen the use of both pattern based as well as model-based methods. Machine learning techniques like neural networks have also been used in sequence classification as in protein sequence classification [85].

iii. Clustering

In this category data is grouped based on their similarity. Clustering is of particular interest in temporal data mining since it provides an attractive mechanism to automatically find some structures in large data sets that would be otherwise difficult to summarize. For example in financial data, it would be of interest to group stocks that exhibit similar trends in price movements. There are variety of methods for clustering sequences, such as model based sequence clustering methods [91, 90, 89], pattern alignment-based scoring [88], similarity measures [87, 86].

iv. Search and Retrieval

Search and retrieval techniques play an important role in temporal data mining as the data to be searched is very huge and the data to be searched is a pattern. The problem is concerned with efficiently locating subsequences in large archives of sequence data. In applications like content-based retrieval approximate matching is done [79]. When the sequence elements are feature vectors, standard metrics such as Euclidean distance may be used for measuring similarity between two elements. Similarity measures based on other transforms have been explored as well. Wu et al [78] present a comparison of Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) based similarity searches. Another approach is to regard two sequences as similar if they have enough non-overlapping time-ordered pairs of subsequences that are similar, which was applied to find matches in US mutual fund database [36].
Keogh & Pazzani [77] used a piece-wise aggregate model for time series to allow faster matching using dynamic time wrapping.

In time series data one of the main concerns is to identify patterns of interest. Unlike in search and retrieval applications, in pattern discovery there is no specific query in hand with which to search the database. A pattern is a local, repeating and interesting structure in the data. Though there are many ways of defining what constitutes a pattern, a typical pattern could be in the form of a substring with some ‘don’t care’ characters in between. A frequent pattern is one that occurs many times in the data.

Study of sequential patterns is an active research topic and has a usage in wide range of applications. Each of these applications has a different model and solution for their use, using sequential patterns. All these models are mainly categorized into four groups viz., frequent patterns, periodic patterns, statistically significant patterns, and approximate patterns. Among these groups our interest lies in frequent patterns.

Methods for finding frequent patterns are considered important because they can be used for discovering useful rules and the rules have been popular representations of knowledge in machine learning and AI for many years. In data mining, association rules are used to capture correlations between different attributes in the data [35]. The conditional probability of the consequent occurring given the antecedent is referred to as confidence of the rule. For example, in a sequential data stream, if the pattern “B follows A” appears $f_1$ times and the pattern “C follows B follows A” appear $f_2$ times, it is possible to infer a temporal association rule “whenever B follows A, C will follow too” with a confidence ($f_1/f_2$). The rule is interesting only if it has high confidence and it is applicable sufficiently often in the data.

**Sequential patterns:** The frame of sequential pattern discovery is essentially an extension of the original association rule mining framework proposed for a database of unordered transaction records [37] which is known as
Apriori algorithm. The sequential pattern mining framework of Agarwal & Srikant [36] which basically extends the frequent item sets idea to the case of patterns with temporal order in them. The temporal patterns of interest are essentially some ordered sequence of item sets. A sequence s of item sets is denoted by \((s_1, s_2, \ldots, s_n)\), where \(s_j\) is an item set. Since s has n item sets, it is called n-sequence. A sequence \(a = (a_1, a_2, \ldots, a_n)\) is said to be contained in another sequence \(b = (b_1, b_2, \ldots, b_m)\) if there exist integers \(i_1 < i_2 < \ldots < i_n\) such that \(a_1 \subset b_{i_1}, a_2 \subset b_{i_2}, \ldots, a_n \subset b_{i_n}\). That is, an n-sequence a is contained in a sequence b if there exists an n-length subsequence in b, in which each item set contains the corresponding item sets of a. Any sequence of item sets with support greater than or equal to the user specified minimum support threshold is called a large sequence. If a sequence is large and maximal then it is regarded as a sequential pattern. The concept of a sequential pattern is quite general and can be used in many other situations.

The search for sequential pattern begins with the discovery of all possible item sets with sufficient support. The support of an item set is the fraction of transaction sequences in which at least one transaction contains the item set. The next step is called the sequence phase, where multiple passes are made over the data, a set of new potentially large sequences called candidate sequences are generated. Two families of algorithms are presented by Agarwal & Srikant [36] and are referred to as count-all and count-some algorithms. The count-all algorithm first counts all the large sequences and then prunes out the non-maximal sequences in a post-processing step. The count-some algorithm intelligently exploits the maximality constraint. The idea is to count longer sequences first, since they anyway contain the other sequences which need not be counted. Thus the count-some algorithms have a forward phase, in which all frequent sequences of certain lengths are found, and then a backward phase, in which all the remaining frequent sequences are discovered.

**Frequent episodes:** The difference between sequential patterns to that of frequent episodes is that in sequential patterns we are given a collection of
sequences and the task is to discover sequences of items that occur in sufficiently many of those sequences, where as in frequent episodes the data are given in a single long sequence and the task is to unearth temporal patterns called episodes that occur sufficiently often along that sequence. Manila et al [41] apply frequent episode discovery for analyzing alarm streams in a telecommunication network. Frequent episode mining is used here as an alarm management system. The goal was to improve understanding of the relationships between different kinds of alarms and thereby provide some early warning about which alarm often goes off close to one another. The framework of frequent episode discovery is to represent a event sequence S denoted by \{(E_1,t_1), (E_2,t_2), \ldots\} where \(E_i\) takes values from a finite set of event types \(E\) and \(t_i\) is an integer denoting the time stamp of the \(i\)th event. The sequence is ordered with respect to the time stamps so that, \(t_i \leq t_{i+1}\) for all \(i=1,2,\ldots\). An episode is just a partially ordered set of event types. When the order among the event types of an episode is total, it is called serial episode and when there is no order at all, the episode is called a parallel episode. One can have episodes that are neither serial nor parallel; the episode discovery framework of Mannila et al. [41] is mainly concerned with only these two varieties of episodes.

### 2.6 Data Mining in Stock Market Predictions

In stock market business, investors want to maximize their returns by buying or selling their investments at an appropriate time. Since stock markets data are highly time-variant and are normally in a nonlinear pattern, predicting the future trends of a stock is a challenging problem. Over the period researchers have chosen different techniques to predict the markets. Let us investigate these techniques and methods.

The task of meeting our objective in identifying better price prediction for a stock, using data mining is met considering the following strategy.
i. Choosing an appropriate data mining technique of classification and prediction for stock market application. Initially we classify the historic data into appropriate patterns and save it as knowledge base for further use.

ii. Handling continuous time series data from the stock market and taking this current data as a reference match the pattern with the existing patterns from the knowledge base.

iii. Considering ‘Elliot Wave Principle’ for identifying significance of the transitional patterns in the data.

For many of the emerging application it is demanding to conduct advanced analysis and data mining over fast and large data streams to capture the trends, patterns and exceptions. Previous studies argue that mining data streams is challenging in two respects. On one hand, random access to fast and large data streams may be impossible. Thus, multi-pass algorithms are often infeasible. On the other hand, the exact answers from data streams are often too expensive. Thus approximate answers are acceptable. The key to these issues is online mining of changes. In general, which the patterns in snapshots of data streams are important and interesting, the changes to the patterns may be more critical and informative [92].

Many of the Data Mining tools were used in predicting different attributes of stock market such as Intraday and Inter-Stock price movements, Stock index movements, financial events and related stock movements etc. Jo Ting et al. in their paper [55] have predicted intraday and inter-stock prediction using pattern associative classification mechanism. In their paper a pattern-based stock data mining approach which transforms the numeric stock data-to symbolic sequences, carries out sequential and non-sequential association analysis and uses the mined rules in classifying and predicting the further price movements is proposed. They have used three different methods for carrying out associative classification and prediction namely, Best Confidence, Maximum Window Size and Majority Voting. They select the mined rules and make the final prediction. Another
important factor in stock market is the timing, when to buy or when to sell at the future price of the stock. Mark O Afolabi et al. in their paper [56] have used new methods such as Fuzzy logic, Neural Network and hybridized methods such as hybrid Kohonen Self Organizing Map (SOM), Adaptive Neuro-Fuzzy Inference System (ANFIS) etc. The methods were used to increase the accuracy of prediction in determining stock market timing, and the future price of a stock. Senthamarai Kannan et al. in their financial stock market forecasting paper have use data analysis as data mining technique for predicting future price movements in terms of its increase or decrease from the current price. Five methods of analyzing stocks were combined to predict if the day’s closing price would increase or decrease. These methods were Typical Price (TP), Bollinger Bands, Relative Strength Index (RSI), CMI and Moving Averages (MA). The paper also discusses various techniques which are able to predict with future closing stock price will increase or decrease better than certain level of significance.

The paper [58] by Paul D Yoo et al. is a survey paper on machine learning techniques for stock market prediction. They have presented different developments in stock market prediction models, and discussed their advantage and disadvantages. Through this survey they found that incorporating event information with prediction model plays very important roles for more accurate prediction. They have concluded that an accurate event weighting method and a stable automated event extraction system are required to provide better performance in financial time series prediction. Associated with events in financial market Anny Ng et al. have presented a paper [59] on mining frequent episodes and the effect of these events on stock prices. They have taken political and economic events from Chinese local newspapers and related it with the movement of stock market prices. They have also proposed a new method for the mining of frequent temporal patterns. Similarly authors Narcyz Roztocki et al. in their paper [60] studied the growing repository of events in the field of information technology research. The empirical results presented indicate that financial markets differentiate among technologies that companies invest in to integrate their information systems. The study also confirms that technology
maturity, financial health of the investing company, and stock market conditions are important factors influencing stock market reaction.

The use of intelligent systems for stock market predictions has been widely established. Several methods were adopted to predict stock market index movements. Previous studies in stock market predictions using artificial intelligence techniques such as artificial neural networks and case-based reasoning have focused mainly on spot market prediction. Kyoung-jae Kim et al. in their paper [62] adopted a mechanism to predict the daily up/down fluctuant direction of the price for Korean Stock market index futures to meet this recent surge of interest. The forecasting methodologies employed in their research are the integration of genetic algorithm and artificial neural network and integration of genetic algorithm and case-based reasoning. In the paper [61] Crina Grosan et al. have introduced a genetic programming technique called Multi-Expression programming for the prediction of two stock indices. The performance of the method is compared with an artificial neural network.

Different methods used for stock market prediction has motivated us to use them in predicting the price of a stock for a day at different intervals using a combination of techniques. We look into patterns in the historical data and try to find some association between these similar patterns to come out with a technique to predict the price of a stock. In the process we also try to look for important episodes in the data and their effect in the movement of the stock market prices.