**2.1 INTRODUCTION**

Prediction of asset prices has always been an interesting area of research

**2.2 MARKET EFFICIENCY**

Ever since the term EMH has come into the financial economics

**2.2.1 Meaning of Efficient Markets**

Prior to Fama et al (1969), the literature on the time series of prices

had increasingly but heuristically connected its random walk models with

competitive markets. They defined efficient market as, a market that adjusts

**2.2.2 Notable Empirical Studies in Market Efficiency**

Samuelson (1965), in his seminal work, Proof That Properly

**2.2.3 Implication from Literature**

The theory of EMH is simple in principle, but remains elusive. The

EMH came to be supported by a growing body of empirical research

demonstrating the difficulty of beating the market, whether by analysing

**2.3 PREDICTABILITY OF STOCK MARKETS**

There is a large body of research carried out suggesting the

predictability of stock markets. Lo & MacKinlay (1988) in their research

paper claim that stock prices do not follow random walks and suggested

considerable evidence toward predictability of stock prices. Basu (1977),

**2.3.1 Data Mining Approaches to Stock Market Prediction**

As the dynamic stock market leaves a trail of huge amount of data,

storing and analyzing these tera bytes of information has always been a

challenging task for the researchers. After the advent of computers that have

brute processing power, storage technologies such as databases,

**2.3.2 Application of ANN, k-NN, Decision Tree, and SVM**

**Techniques in Stock Market Prediction and Classification**

Hertz et al (1991) offer a comprehensive view of neural networks

and issues of their comparison to statistics. Hinton (1992) investigates the

statistical aspects of neural networks. Weiss & Kulikowski (1991) offer an

account of the classification methods of many different neural and statistical

models.

26

Brock et al (1992) found nonlinearities in market prices and

showed that the use of technical analysis indicators, under certain

assumptions, may generate efficient trading rules (Brock et al 1992). The

prices of other financial commodities also have this nonlinear dynamic

property. Savit (1989), for example, suggested a nonlinear dynamic model for

option prices.

Kuan & White 1994; Bierens 1994; Lewbel 1994, The main focus

on the ANN technology, in application to the financial and economic fields,

has so far pertaining to non-linear relationship among variables . Many

economists advocate the application of neural networks in different fields of

economics (Kuan & White 1994; Bierens 1994; Lewbel 1994). According to

**2.3.3 Inference from Literature**

The review of various works brings out various interesting facts.

There are many studies that applied data mining tools to investigate the

predictability of various financial series. The reasons quoted for applying

data mining tools are its ability to handle voluminous data and nontrivial

extraction of implicit, previously unknown, and potentially useful information

from financial data.

**2.3.4 Application of Data Mining Tools for Studying Association**

**among Stock Indices**

Liao et al (2001) investigated the co-movement between foreign

exchange rates and Taiwan stock indices. They implemented the association

ruling approach to explore the co-movement between foreign exchange rates

and category stock indices in Taiwan. Data of Forex rates and stock indexes

were collected and the Apriori algorithm was used to generate the association

rules. The study proposed several possible portfolio alternatives in the Taiwan

financial capital market including foreign exchange currencies and stock

investment under different circumstances.

**2.3.5 Inference from Review**

The review of various articles that employed various statistical and

data mining association rule mining techniques showed that understanding the

association among various financial time series, especially the stock market

indices is a fascinating area of research. The review of past studies indicates

that researchers aim at studying the influence and relationship between the

stock prices and various macroeconomic variables and technical indicators. It

is notable that not many research attempts are made to study the influence of

global cues (gold prices, crude oil prices, LIBOR rate, leading stock indices,

etc) on the movement of a particular stock index price. Research studies

employ the cross correlation techniques and the Apriori algorithm popularly

to study the association among various stock price indices. The Indian stock

**2.4 DATA AND SOURCES OF DATA**

**2.4.1 Stock Market Indices**

Stock market indices are meant to capture the overall behavior of

equity markets. A stock market index is created by selecting a group of stocks

that are representative of the whole market or a specified sector of the market.

A stock index is calculated with reference to a base period and a base index

value.

**2.5 RESEARCH TOOLS**

**2.5.1 k-Nearest Neighbor Algorithm**

Nearest Neighbors is a data mining prediction technique that looks

for records with similar predictor values in the historical database and uses

their prediction values for making prediction for an unclassified record. As

the name implies, this algorithm looks for k nearest cases in sample dataset

and uses the outputs of those nearest cases for forecasting the output for a new

instance of data. All neighbors output can be given equal weights or the

closest neighbors can be given more weights which is inversely proportional

to its distance from the new data record. One of the most widely used metrics

for identifying the nearest neighbor is Euclidean distance**.**

**2.5.2 Artificial Neural Network**

An artificial neural network is an efficient information processing

system that resembles the function of human brain. An ANN contains several

nodes with an input layer and an output layer. Each node emulates a neuron as

in the brain and is connected to the other by a connection link. Each

**2.5.3 Support Vector Machines**

The SVM is a promising new machine learning algorithm that uses

nonlinear mapping to transform original training data into higher dimensional

plane so that a demarcating plane can be found between class labels. It then

employs optimization techniques to optimize the width of this hyperplane. By

identifying support vectors, which are instances touching the boundary of the

marginal hyper plane, a new instance can be easily classified. By modeling

the input-output relationship using sequential marginal optimization

regression technique, the SVM can be adapted also for forecasting.

**2.5.4 Decision Trees**

Decision trees are represented by a set of questions that splits the

entire dataset into smaller and smaller subgroups. It searches for one predictor

variable and its particular value that splits the entire dataset into two parts

with maximum homogeneity in terms of decision variable. The choice of split

variable is based on impurity function. This splitting process is then repeated

with each of the resulting data fragments until leaves or decision nodes are

reached. The resulting tree can be used for making decision rules for

prediction. A powerful ensemble variant of decision tree called random forest

is used as a classifier in this study.

**2.5.5 Apriori Algorithm**

Apriori is a classic algorithm proposed by R. Agarwal and R.

Srikant in 1994 for mining frequent item sets in transactional databases. The

name of the algorithm is based on the fact that it uses prior knowledge of the

frequent item set properties. It proceeds by identifying the frequent individual

items in the database that satisfy minimum support level. It then employs an

iterative approach by extending them to larger and larger item sets until no

51

more frequent item sets are found in the database. The frequent item sets

determined by Apriori lead

**2.7 CONCLUSION**

The review of various literatures in the field of financial forecasting

and prediction revealed that numerous interesting studies are being carried out

to understand the phenomena of predictability of financial markets. The

notable work of Eugene Fama paved as a stepping stone that raised interesting

queries questioning the predictability to stock market movements. From then