Volume 4, Issue 1, January 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Forecasting the Egyptian's Market Trend Using Rough Set

Naglaa Ragaa Saeid Hassan
Department Of Operations Research, Institute of Statistical Studies and Researches
Cairo University - Egypt

Abstract*— Most researchers have used artificial intelligence (AI) techniques to predict the price of every trading day, week, and month. However, it is more important to determine stock market best time; when to buy and when to sell stocks, than to predict the price movement for everyday while investors in stock market generally do not trade everyday. If investors trade their stocks everyday, they would be charged a tremendous amount of trade fees. Besides, the decision to invest is initially influenced by trend and then by the prices. In this paper the aggregate purchases and sales of equities are divided among five major investment groups in the economy: Egyptian, Arabian, foreigners, individuals and institutions. Different investors groups present different investment behaviour and of course their trades may relate to asset prices in different manners. This study examines whether some investors groups are associated with concurrent or subsequent market-wide price movements through the market trend using rough set theory as one of an artificial intelligence techniques. The experimental results of this study prove on the following: (1) The ability of the rough set approach to discover facts hidden in data which represent the market trend. (2) The efficiency of the suggested classifying method to manage the discretization reaching to the general market trend. (3) Using the generalization to interpret the extracting trading rules which represent the market trend to proof with self evident on general known trading rules such as in case of selling the price is downward trend in the market and vice versa.

***Keywords*— Egyptian's market trend, Rough set, Trading rule, Investment groups, EXG30**

I. INTRODUCTION

Large number of researchers is undertaking the behaviour study of stock markets all round the world. Each researcher work throws new insight owing to the dynamic nature of the capital market. There is never end for that going on debate to tell whether the stock markets are predictable or not [2]. Many research studies have shown empirical evidence stating that the stock markets are not predictable and there is equal number of research works showing the predictability of stock market prices. Before the age of computers, people traded stocks and commodities depending primarily on intuition. As the level of trading grew, people searched for tools that would increase their gains while minimizing their risk. Technical analysis, statistics, fundamental analysis, and linear regression are all used to attempt to predict the market's direction. None of these techniques has proven to be the consistently correct prediction tool that is desired and many analysts argue about the usefulness of many of the approaches.

Traders often use technical tools such as Moving Average (MA), Momentum (M), and Stochastic to guide their investment decisions. They often rely on more than one method to predict the future price of shares in the market. The main assumption behind the use of technical analysis is that history repeats itself, and therefore by identifying the recurring patterns, future price movements are predictable.

In the recent years, investors have started to show interest in trading on stock market indices as it provides an opportunity to hedge their market risk, and at the same time it offers a good investment opportunity for speculators and arbitrageurs. Where, the decision to invest is initially influenced by the trend and then by the prices. Investors adopt unique trading strategies depending on the trend prevailing in the stock market. Thus, the possible trend that would prevail in the stock market, could be helpful for investors to adopt a suitable investment strategy as in [2].

Typically stock market indices are the performance indicators for the entire market in general. Interested investors all around the globe keep track of the movement of various stock market indices to get an idea about how the global markets are moving. Thus, in the stock market, financial investment decision making requires the accurate prediction of the future states based on the identification of hidden patterns in the available historical data.

This paper focuses on recognizing the market trend using rough set theory within a generalization concept of generating trading rules for market timing. Different investors groups present different investment behaviour and of course their trades may relate to asset prices in different manners. This paper analysis the relation between stock market INDEX EGX 30 and cash flows from a broad array of investor groups (Egyptian, Arabian, foreigners, individuals and institutions) using rough set as one of an artificial intelligence techniques. So, the stock market's movements are predicted in order to retrieve knowledge that could guide investors on when to buy and sell. The experimental results of this study shows that rough set able to discover important facts hidden in data so, obtaining the market trend while eliminating superfluous factors in noisy stock market data to extract these trading rules.

II. METHODOLOGY

A. Rough set: Foundations

Rough sets theory was introduced by Pawlak in 1982. It was developed based on mathematical tool to deal with vagueness and uncertainty in the classification of objects in a set as in [3]. The rough set philosophy is founded on the assumption that there is some information regarding features which can be associated with every object of the universe. In rough sets, the data is organized in a table called decision table, which are flat tables containing attributes as columns and data elements as rows. The class label is called as decision attribute, the rest of the attributes are the condition attributes. For rows, rough set theory employs the notion of indiscernible class to group similar topless together; while for columns, it employs the notion of indiscernible attribute to identify significant attributes as in [4].

We call the class label a decision feature, the rest of the features are conditional. Let O, F denote a set of sample objects and a set of functions representing object features, respectively. Assume that $B \subseteq F, x \in O.$ Further, let $[x]_B$ denote:

$$[x]_B = \{y : x \sim B\ y\}$$

Rough set theory define three regions based on the equivalent classes induced by the feature values: lower approximation $\underline{B}X$, upper approximation $\overline{B}X$, and boundary $BND_B(X)$. A lower approximation of a set X contains all equivalence classes $[x]_B$ that are subsets of X, and upper approximation contains all equivalence classes $[x]_B$ that have objects in common with X, while the boundary is the set of all objects in $\overline{B}X$ that are not contained in $\underline{B}X$. So, we can define a rough set as any set with a non-empty boundary as in [5]. The indiscernibility relation ~ B (or by Ind_B) is a fundamental principle of rough set theory. Informally, ~ B is a set of all objects that have matching descriptions. Based on the selection of B, ~B is an equivalence relation partitions a set of objects O into equivalence classes. The set of all classes in a partition is denoted by O/~ B (also by O/ Ind_B). the set O/ Ind_B is called the quotient set. Affinities between objects of interest in the set $X \subseteq O$ and classes in a partition can be discovered by identifying those classes that have objects in common with X. Approximation of the set X begins by determining which elementary sets $[X]_B \in O/\sim B$ are subsets of X. For a detailed review of the basic material, reader may consult sources such as in [3, 6-9].

B. Prediction model using rough sets:

As in [10], functionally, the proposed Rough Set Prediction Model (RSPM) of stock market data can be partitioned into four distinct phases:

- **Data collection phase**
- **Pre-processing phase**: This phase includes tasks such as extra variables addition and computation, decision classes, assignments, data cleansing, completeness, correctness, attribute creation, attribute selection and discretization.
- **Applying rough set model for Analysis and Rule Generating Phase**: This phase includes the generation of preliminary knowledge, such as computation of object reducts from data, derivation of rules from reducts, rule evaluation and prediction processes.
- **Classification and Prediction phase and validation tests**: This phase utilize the rules generated from the previous phase to predict the stock price movement.

III. TRADING RULES EXTRACTION AND THE MARKET TREND

Case study

A. Data collection phase:

The primary source of data in this paper is from "The Egyptian's company for information publishing" which it reports purchases, sells and net of major assets by investor groups in the Egyptian economy. The data covers approximate one year time period daily; from 12 Aug.2008 to 15 Oct. 2009. This study use total purchases and sells of equities for five major investors groups: Egyptian investors, Arab investors, foreign investors, institutions investors and individual investors besides, the data also include the daily opening, closing index values of stock markets and the difference between close and open EGX

B. Pre-processing phase:

In this phase, the decision table required for rough set analysis is created. In doing so, a number of data preparation tasks such as data conversion, data cleansing, data completion checks, conditional attribute creation, decision attribute generation, discretization of attributes are performed. Data splitting is also performed which created two randomly generated subsets, one subset for analysis containing %75 of the objects in the data set and one validation containing the remainder %25 of the objects. It must be emphasized that data conversion performed on the initial data must generate a form in which specific rough set tools can be applied.

Often, real world data contain missing values. Since rough set classification involves mining for rules from the data, objects with missing values in the data set may have undesirable effects on the rules that are constructed. The aim of the data completion procedure is to remove all objects that have one or more missing values. Incomplete data or information systems exist broadly in practical data analysis, and approaches to complete the incomplete information system through various completion methods in the pre-processing stage are normal in data mining and knowledge discovery. However, these methods may result in distorting the original data and knowledge, and can even render the original data to be unable. In this paper, it is intend to delete some data from original data set to study the effect of that deleted data on extracting knowledge from the rules and generality.

To overcome the shortcomings inherent in the traditional methods, we eliminate all incomplete information system (i.e. decision table).

B.1. Data description:

The decision table of this study contains (216 Objects, 18 Attributes, 15 Conditional attribute and 3 decision attributes) refer to as data set 1 (DS1) and because of there often exist conditional attributes that don't provide (almost) any additional information about the objects. These attributes need to be removed in order to reduce the complexity.

B.2. Data selection:

The data selection depending on the optimality criterion associated with attributes of DS1 which affects the final generated rules because the rough sets model can only extract the knowledge inherent with existing data set or table, which it means that if the attributes are not represented in the related information system, then the generated rules would be dispersed and become insignificant. After deep looking inside the data contents of DS1, it found that it is possible to work only with the net attributes as it summarizes the total effect of buyers and sellers. Applying this action of selection on DS1, obtaining a new decision table DS2 which contains (216 Objects, 6 Attributes, 5 Conditional attribute and 1 decision attribute). It is also noticed that according to a relation between the conditional attributes, which:

Egyptians + Arabs + Foreigners = Individuals + Institutions

So, DS2 could be split into two decisions tables and study each one alone as in (TABLE I-1 and TABLE I-2).

TABLE I-1

DESCRIPTION OF DATA (PART I)

Item	**Type**	**Description**
Date	The universe of	Date of transaction
EgyptianN	Conditional attribute	Net number of Egyptian buyers and
ArabN	Conditional attribute	Net number of Arab buyers and sellers
ForeignerN	Conditional attribute	Net number of foreigner buyers and
EGX30D	Decision attribute	Difference between close and open EGX

TABLE I-2

DESCRIPTION OF DATA (PART II)

Item	**Type**	**Description**
Date	The universe of objects	Date of transaction
Individuals	Conditional attribute	Net number of individual buyers and
Institutions	Conditional attribute	Net number of institution buyers and
EGX30D	Decision attribute	Difference between close and open EGX

Notice that:

The TABLE I-1 is containing (216 Objects, 4 Attributes, 3 Conditional attribute and 1 decision attribute) and TABLE I-2 is containing (216 Objects, 3 Attributes, 2 Conditional attribute and 1 decision attribute).

As it is mentioned before, in the Pre-processing phase, the reliability data should be considered where the objects with missing or wrong value will be deleted according to the following relation where:

(Egyptians + Arabs + Foreigners = Individuals + Institutions)

Thus, the raw data set of this project is DS2 and after eliminating the missing and wrong data which includes 132 wrong data objects out of the original 216 objects, So, the rest objects appear in the new decision tables (TABLE I-1 and TABLE I-2) which contains only 84 objects.

B.3. Discretization:

In rough sets theory and other induction learning systems, discretization is an important algorithm and can be viewed as a process of information generalization (or abstraction) and data reduction as in [9]. Discretization is a process of grouping the values of the attributes in intervals in such a way that the knowledge content or the discernibility is not **lost** as in [11-12]. Many discretization approaches have been developed so far. Nguyen S. H had given some detailed description about discretization in rough set as in [3].

Before applying analysis on data set using rough set software or package, first, it must be discretized. As a result of discretization, the precision of the original data will be decreased but its generality will be increased. Of course, the efficiency of the prediction process is increased if applying a suitable discertization algorithm on data set

The discretization of real value attributes is one of the important problems to be solved in data mining, especially in rough set. We know, when the value set of any attribute in a decision table is a continuous value or a real number, and then it is likely that there will be few objects that will have the same value of the corresponding attributes. In such a situation the number of equivalence classes based on that attribute will be large and there will be very few elements in each of such equivalence class, which it will lead to the generation of a large number of rules in the classification of rough set, thereby making rough set theoretic classifiers inefficient as in [11]. As the major objective of stock market analyses is to determine whether the stock market trend is upward, downward or hold. In this study, It is used the values of attribute (EGX 30D) to be (1) if increasing, (-1) if decreasing and (0) otherwise. The same process is applied on all attributes (EgyptianN, ArabN and ForeignersN) in TABLE I-1 and all attributes (IndividualsN and InistitutionsN) in TABLE I- 2.

C. Applying rough sets model

After the pre-processing step, the rough sets model is applied as analysis tool to extract rules from decision tables to predict the stock market trend. In validation step, the rules are tested for their validation. The tested sample is selected randomly and classified only once, with its size being set arbitrarily to a constant percentage (25%) of all objects.

D. Results, analysis and interpretation

Applying rough set software (packages), the rule derived from re-ducts can be used to classify the data. The set of rules is referred to as a classifier and can be used to classify new and unseen data.

TABLE I-1 includes {EgyptianN, ArabN, ForeignersN} and after applying the reduction action, it is generated 9 rules which only 2 are consistent (rules 4 and 9) as shown in Fig 1.

	Rule
1	EgyptianN(-1) AND ArabN(1) AND ForeignersN(-1) => EGX 30D(-1) OR EGX 30D(1)
2	EgyptianN(1) AND ArabN(-1) AND ForeignersN(-1) => EGX 30D(-1) OR EGX 30D(1)
3	EgyptianN(1) AND ArabN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
4	EgyptianN(-1) AND ArabN(-1) AND ForeignersN(-1) => EGX 30D(-1)
5	EgyptianN(-1) AND ArabN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
6	EgyptianN(1) AND ArabN(1) AND ForeignersN(-1) => EGX 30D(1) OR EGX 30D(-1)
7	EgyptianN(0) AND ArabN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
8	EgyptianN(-1) AND ArabN(1) AND ForeignersN(1) => EGX 30D(1) OR EGX 30D(-1)
9	EgyptianN(1) AND ArabN(-1) AND ForeignersN(0) => EGX 30D(1)

Fig 1 The rules generate from re-duct

The rules are generated from re-duct using ROSETTA software contains the following knowledge:

Rule 4: EgyptianN(-1) AND ArabN(-1) AND ForeignersN(-1) => EGX 30D(-1)

- Which means "IF Egyptians, Arabs and foreigners are sellers Then the market trend is going downward"

Rule 9: EgyptianN(1) AND ArabN(-1) AND ForeignersN(0) => EGX 30D(1)

- Which means "IF Egyptians are buyers, Arabs are sellers and foreigners are even Then the market trend is going upward" Otherwise (rules 1, 2, 3, 5, 6, 7 and 8) the market trend may go upward or downward.

So, the TABLE I-1 shows a set of the generated rules. These obtained rules are used to build the prediction system for investors.

- After splitting TABLE II-2 into learning (75%) and validation (25%)

The results of learning, obtaining only one reduct where it is including (EgyptianN, ForeignersN) but attribute ArabN is dispensable.

The reduct generated only six rules; two of them are consistent (rules 1 and 6) as in Fig 2**.**

	Rule
1	EgyptianN(-1) AND ForeignersN(-1) => EGX 30D(-1)
2	EgyptianN(1) AND ForeignersN(-1) => EGX 30D(-1) OR EGX 30D(1)
3	EgyptianN(1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
4	EgyptianN(-1) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
5	EgyptianN(0) AND ForeignersN(1) => EGX 30D(-1) OR EGX 30D(1)
6	EgyptianN(1) AND ForeignersN(0) => EGX 30D(1)

Fig 2 The rules generate from reduct

Notices:

Rule 1: EgyptianN(-1) AND ForeignersN(-1) => EGX 30D(-1)

- Which means that "IF Egyptians are sellers and foreigners are sellers Then the market trend is going downward"

Rule 6: EgyptianN(1) AND ForeignersN(0) => EGX 30D(1)

- Which means "IF Egyptians are buyers and foreigners are even Then the market trend is going upward"
- Otherwise (rules 2, 3, 4, and 5) the market trend may go upward or downward.

- When tested on validation set, the accuracy is equal to 0.714286.

Comment: The results are compared with the actual observed trend, and it is seen that the rough set model has predicted the trend accurately. Whereas, if the accuracy is higher, the classification is less ambiguous and the quality of rule is better, therefore, the classification is unambiguous and acceptable

From TABLE I-2, it is obtained only 1 reduct including (IndividualsN, InstitutionsN) and generated 3 rules from which only 1 is consistent (rule 3) as in Fig 3 and the same result in learning

	Rule
1	IndividualsN(-1) AND InstitutionsN(1) => EGX 30D(-1) OR EGX 30D(1)
2	IndividualsN(1) AND InstitutionsN(-1) => EGX 30D(-1) OR EGX 30D(1)
3	IndividualsN(-1) AND InstitutionsN(-1) => EGX 30D(-1)

Fig 3. The rules generate as reduct

Notice that:

Rule 3: IndividualsN(-1) AND InstitutionsN(-1) => EGX 30D(-1)

- Which means "IF Individuals are sellers and Institutions are sellers Then the market trend is going downward"
- Otherwise (rules 1 and 2) the market trend may go upward or downward.

- When tested on validation set the accuracy is equal to 0.761905.

E. Prediction phase

In this phase, the generated rules are utilized to predict the market trend. Combining all previous results, it is noticed that:

1- The attribute ArabN is dispensable.

2- The stock market goes upward only IF Egyptians are buyers and foreigners are even

3- The stock market goes downward

IF Egyptians are sellers and foreigners are sellers

Or/ and

IF Individuals are sellers and Institutions are sellers

And this rule as proof with self evident on general known trading rules such as in case of selling the price is downward trend in the market and vice versa

4- Different investors groups present different investment behaviour and of course their trades may relate to asset prices in different manners.

5- Some investors groups are associated with concurrent or subsequent market-wide price movements through the market trend.

6- So, the market trend is predicted using the rough set model applying 1 year data. When the results are compared with the actual data and observed trend, it is seen that the rough set model has predicted the market trend with high accuracy. So, these extracting trading rules could be used to guide investors whether to buy, sell or hold on the stock

V. CONCLUSIONS

Different investors groups present different investment behaviour and of course their trades may relate to asset prices in different manners. In this paper, the relation between stock market INDEX EGX 30 and cash flows from a broad array of investor groups (Egyptian, Arabian, foreigners, individuals and institutions) is analyzed using rough set as one of the artificial intelligence techniques. This paper proves on the following: First, the ability of the rough set approach to discover facts hidden in data which represent the market trend as an index. Second, this study is a self evident on general known trading rules such as in case of selling the stocks, the price is downward trend of the market and vice versa by generalization the extracting trading rules representing the market trend. It has been demonstrated that suggested model is a successful and promising forecasting tool for predicting the general future trend in the market where the experimental results are encouraging and prove that the usefulness of suggested method obtaining a higher strength of prediction in terms of the significance of rules which means it could be used to guide investors whether to buy, sell or hold a stock. In spite of eliminating the missing values, the internal dependency structure of the system is kept intact. Besides, it is showed that the efficiency of the suggested classifying method to manage the discretization reaching to the general trend.

Thus, rough set is an important tool to deal with uncertain or vague where it helps the researchers in identifying the future trends in the markets. Besides, it is a good indicator for the investors or the decision makers to make better investment decisions.

REFERENCES

[1] R. Lawrence “Using Neural Networks To Forecast Stock Market Prices”, Department of computer Science, University of Manitobo, pp. 5-10., 1997 Retrieved 2006 from: http://www.cs.uiowa.edu/~rlawrenc/research/Papers/nn. pdf.

[2] S. Vasanthi, M. V. Subha & S. Th. Nambi, “An empirical study on stock index trend prediction using Markov chain analysis” Sri Krishna International Research & Educational Consortium, 2011, JBFSIR V. 1, I. 1.

[3] Z. Pawlak, B. J. Grzymala, R. Slowinski and W. Ziarko, “Rough sets”, Communications of the ACM, V. 38, No. 11, pp. 89-95, 1995.

[4] S. H. Jaaman, S. M. Shamsuddin, B. Yusob and M. Ismail, “A predictive for Malaysian stock market returns”, International Research Journal of Finance and Economics Euro Journal publishing, I. 30, 2009.

[5] H. Al-Qaheri, A. E. Hassanien and A. Abraham, “Discovering Stock Price Prediction Rules Using Rough Sets”. Neural Network World Journal, 2008, (to appear).

[6] R.H. Golan and W. Ziarko “A methodology for stock market analysis utilizing rough set theory”, Proc. of the IEEE/IAFE Computational Intelligence for Financial Engineering, 9-11 Apr, pp 32-40, 1995.

[7] Z. Pawlak, “Rough Sets”, Int. J. of information and computer science., V. 11 p. 341-356, 1982.

[8] Z. Pawlak, “Rough Sets”, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Netherlands, 1991.

[9] L. Polkowski, “Rough Sets: Mathematical Foundations”, Physica-Verlag, 2003.

[10] F. E.H. Tay and L. Shen, "Economic and financial prediction using rough sets model Computing", European Journal of OR, 141, p 641–659, 2002.

[11] H. S. Nguyen and A. Skowron, “Quantization of real value attributes”, Proc. of Second Joint Annual Conf. on Information Science, Wrightsville Beach,North Carofina, p34-37, 1995.

[12] C. Y. Chen, Z. Guoli, S. Y. Qiao and S. P. Wen, “Study on discertization in rough set based on genetic algorithm”, IEEE Proc. of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2003.

