A Survey of Application of Data Mining Techniques in Stock Markets

Mervat M. Ramadan

Professor of Statistics, Statistics, Mathematics, and Insurance Faculty of Business, Benha University

Dina S. Eltelbany

Lecturer of Statistics, Statistics, Mathematics, and Insurance Faculty of Business, Benha University

Mohab Soliman

Teaching Assistant Statistics, Mathematics, and Insurance, Faculty of Business, Benha University

A Survey of Application of Data Mining Techniques in Stock Markets

Mervat M. Ramadan
Professor of Statistics, Statistics, Mathematics, and Insurance
Faculty of Business, Benha University

Dina S. Eltelbany
Lecturer of Statistics, Statistics, Mathematics, and Insurance
Faculty of Business, Benha University

Mohab Soliman
Teaching Assistant Statistics, Mathematics, and Insurance,
Faculty of Business, Benha University

Abstract

Finding effective ways to summarise and visualize stock market data so that people or institutions can use it to make investment decisions is one of the most significant problems in modern finance. The stock market generates a massive amount of valuable data, which has drawn researchers to investigate this problem domain using various methodologies. Long-term, intensive research on these issues was motivated by significant potential benefits. The significance of its applications and the growing information generation have made data mining research very appealing. An overview of the use of data mining techniques, including regression and neural networks, in the stock market, is given in this paper.

Keywords: *Data Mining, Stock Market, Regression, and Neural Network.*

1. Introduction

Financial markets are a complex, nonlinear, non-parametric system, which is extremely difficult to model reasonably (OU &Wang 2009). The significant factors comprise economic condition (e.g., interest rates, exchange rates, monetary growth); psychological investors' variables (e.g., investors' expectations and institutional investors' choices); political variables (e.g., the occurrence and release of critical political events) and other unexpected events (Enke & Thawaornwong 2005); (Das et al. 2017)]. Therefore, investors have been trying to find a way to predict stock prices by choosing the correct and profitable stock within the right time to buy or sell. To achieve those objectives, researchers (Al-Debie & Walker 1999) use fundamental analysis techniques. In contrast, trading rules are developed based on the information associated with macroeconomics, industry, and company-specific variables. In response to such difficulty, data mining techniques, like decision trees, rough set approach, and artificial neural networks, have been introduced for this financial prediction. Data mining is obtaining helpful information to warehouse an extensive database. Data mining can also be interpreted as extracting new knowledge from comprehensive data, which helps in decision-making and discovering an exciting pattern. Some Data mining functionalities can be construed as learning the concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis (Al-Radaideh, 2013).

Consequently, Data mining can be known as knowledge discovery in databases (KDD). Data mining has actively applied to the stock exchange since the 1980s (Sherdiwala, 2014). Data mining techniques have been presented for the prediction of movement signs of the stock market index since the results of [(Leung

et al. 2000) and (Chen et al. 2003)], three where Linear Discriminant Analysis (LDA), Logit and Probit and Neural Network were proposed and compared with parametric models such as GMM-Kalman filter. Kim (2003) applied new and powerful data mining techniques, Support Vector Machines (SVM) and neural network to forecast the direction of stock index prices based on economic indicators. Kumar & Thenmozhi (2006) collected five approaches: SVM, random forecast, neural network, logit and LDA, to predict Indian stock index movement based on variable economic indicators. Ou & Wang (2009) utilized ten different data mining techniques to predict the Hong Kong stock market price movement of the Hang Seng index. The main objective of our study is to build a model to help investors decide the best time to buy or sell stocks based on the knowledge extracted from the historical prices of proposed stocks. The process of decision-making will be based on data mining techniques. The general model of stock price movement for Egyptian exchange markets (EGX) in the line of (Ou and Wang 2009) is described as follows:

$$\boldsymbol{D_t = f(O_{t-1}, H_{t-1}, L_{t-1}, EGX_{t-1}, F_{t-1})} \qquad \textbf{(1)}$$

Where;

- $\boldsymbol{O_{t-1}}$ Denotes the Open Price of the EGX index at time *t-1*.
- $\boldsymbol{H_{t-1}}$ It is the high Price of the EGX index at time *t-1*.
- $\boldsymbol{L_{t-1}}$ It is the low Price in a day of the EGX index at time *t-1*.
- $\boldsymbol{EGX_{t-1}}$ It is the closing price of the EGX (30,70,100) index at time *t-1*.
- $\boldsymbol{F_{t-1}}$ It is the currency exchange rate between L.E and US dollar.

- $\boldsymbol{D_t}$ The direction EGX movement at time t is defined as total value "1" if the closing price at *t* is greater than the closing price at time *t-1*, and "0" otherwise. This function can be linear or nonlinear and will be estimated using data mining algorithms.

2. Data Mining Techniques

Data mining is an analytical process intended to examine data (typically large amounts of data, typically about business or the market) in search of recurrent patterns and systematic relationships between variables and then validate the findings by applying the detected patterns to new subsets of data. Predictive data mining is the most popular and has the most practical business applications because prediction is the goal of data mining. Three stages make up the data mining process:

1) The initial exploration.
2) The model building or pattern identification.
3) The application of the model to new data to generate predictions.

2.1 Descriptive Data Mining

Association Rules Mining: this is an approach for exploring the relationships of interest between variables in massive databases. Examples of association rules mining techniques are Frequent Pattern (FP) Growth and Apriori.

Clustering is the technique of collecting similar objects which are far different from the other groups' things.

Anomaly Detection: this technique is responsible for detecting outliers.

Rough Sets Analysis: this analysis is mainly concerned with analyzing uncertainty and incomplete information.

2.2 Predictive Data Mining

The predictive task uses specific variables or values in the data set to predict future values of other variables of interest.

Classification: such as Decision tree-based methods, neural networks, support vector machines (SVM), naive Bayes classifier, and k-nearest neighbour (KNN) are methods that have been used in solving real-world problems.

Regression: for example, Generalized Linear Model (GLM) performs linear regression for continuous target values, in which the dependent variable is constant, and the independent variable can be straight or discrete.

Classifier Ensembles: present the concept of aggregating multiple classifiers as a novel approach to improve the performance of classifiers that work individually.

3. Application of Data Mining Techniques in the Stock Market

Discovering market trends, creating investment plans, choosing the right stocks, and determining the best time to buy them are all part of stock market forecasting. Financial institutions generate enormous data sets that serve as a foundation for using data mining tools to approach these incredibly complex and dynamic problems. Long-term, intensive research on these issues was motivated by significant potential benefits.

3.1Application of Regression Techniques in the Stock Market

Utilizing data to ascertain their relationship is called data analysis. Tools for modelling and measuring relationships include statistical regression. Several regression techniques have been developed to examine and understand the Big Data model. Simple linear regression analysis is frequently used in various data analysis fields, particularly machine learning, chemistry, and biology. The process of estimating relationships between variables is known as regression analysis. This approach uses various modelling and analysis techniques to examine particular and

distinctive variables, concentrating on the correlation between the dependent variable and one or more independent variables. In machine learning, regression is frequently used to forecast how one variable will behave about the value of another variable. Many factors work together to create or cause an event or situation. All elements are imagined together, and these items are predictive variables. Regression is the mathematical method of discovering the relationship between two or more variables in artificial intelligence.

Bharatiya et al. (2017) applied linear regression to forecast the data set. They demonstrated that their suggested method outperformed all other regression techniques, and stockholders can use that information to make personal investments. Finally, Alsayed et al. (2023) applied several statistical methods, namely ridge, lasso, principal components, and partial least squares (PLS) regression versus elastic-net regression based on empirical mode decomposition, which can overcome the non-stationary problem and nonlinearity characteristics.

3.2Application of Neural Network Techniques in the Stock Market

Some examples of neural network applications in stock markets are presented in this section. Nowadays, it is a widely held belief that enormous sums of money are traded on stock exchanges all over the world. The health of stock markets has a significant impact on the performance of national economies. Additionally, the markets have recently improved as a tool for investment for both ordinary and strategic investors. As a result, they are more closely linked to microeconomic factors and have a more significant direct impact on daily life. Therefore, they make up a mechanism with substantial and immediate social effects. The uncertainty, which is related to their short- and long-term future state, is a feature shared by all stock markets. When the stock market is chosen as the investment tool, this feature

is undesirable for the investor but also inevitable. The best one can hope for is to try to lessen this uncertainty. Predictions about the stock market are one tool used in this process. The main benefit of neural networks is their ability, with the correct number of hidden units, to approximate any nonlinear function to any degree of accuracy.

The range of options available to investors has increased with the development of effective communication and trading facilities. Enke and Thawornwong developed an information gain technique for data mining and machine learning to assess the predictive relationships between various financial and economic variables. The effectiveness of neural network models for level estimation and classification in making future value forecasts is then investigated. Additionally, a cross-validation technique is used to enhance the generalizability of several models. The findings demonstrate that trading strategies guided by classification models outperform buy-and-hold methods and those recommended by level-estimation-based forecasts from neural network and linear regression models regarding risk-adjusted profits (Enke & Thawornwong, 2005).

Defu et al. (2007) discussed using multilayer backpropagation (BP) neural networks, a well-known neural network technique, in financial data mining. An intelligent mining system is created, and a modified neural network forecasting model is presented. The system helps stock investors make decisions by predicting the buying and selling signals based on expected future trends in the stock market. Using neural networks to forecast financial time series is advantageous so that various investors can profit from it. This is shown by the simulation result of seven years to the Shanghai Composite Index, which demonstrates that the return achieved by this mining system is approximately three times as large as that achieved by the buy-and-hold strategy.

4. Conclusion

Financial data are being created and amassed at a rate that has never been seen before, thanks to the expansion of economic globalization and the development of information technology. As a result, there has been a critical need for automated approaches to effectively and efficiently utilise massive amounts of financial data to support businesses and individuals in their strategic planning and investment decision-making. Data mining techniques have been used to find hidden patterns and forecast upcoming trends and behaviours in the financial markets. Increased revenue, lower costs, and significantly better market responsiveness and awareness are just a few of the competitive advantages data mining brings. As a result, much study and practice have been devoted to investigating data mining methods for resolving financial issues.

References

Al-Debie, M., Walker, M. (1999). Fundamental information analysis: An extension and UK evidence. *Journal of Accounting Research*, *31*(3), 261–280.

Al-Radaideh, Q. A., Assaf, A. A., & Alnagi, E. (2013). Predicting stock prices using data mining techniques. *In The International Arab Conference on Information Technology (ACIT'2013),* 17–19 December 2013. Katumu, Sudan

Alsayed, A., Ariç, K. H., & Sek, S. K. (2023). The Behavior of Stock Market Index During the Coronavirus Pandemic in Turkey. *Panoeconomicus*, 1-14.

Bhuriya, D., Kaushal, G., Sharma, A., & Singh, U. (2017, April). Stock market predication using a linear regression. In *2017 international conference of electronics, communication and aerospace technology (ICECA)* (Vol. 2, pp. 510-513). IEEE.

Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, *30*(6), 901-923.

Das, D., Sadiq, A. S., Ahmad, N. B., & Lloret, J. (2017). Stock Market Prediction with Big Data Through Hybridization of Data Mining and Optimized Neural Network Techniques. *Journal of Multiple-Valued Logic & Soft Computing*, *29*, 157-181.

Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with applications, 29*(4), 927-940.

Kim, K.J. (2003). Financial time series forecasting using support vector machines, *Neuralcomputing, 55*, 307-319.

Kumar, M., & Thenmozhi, M. (2006, January). Forecasting stock index movement: A comparison of support vector machines and random forest. In *Indian institute of capital markets 9th capital markets conference paper*.

Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of forecasting*, *16*(2), 173-190.

Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science*, *3*(12), 28-42.

Sherdiwala, K. B. (2014). Data Mining Techniques in Stock Market. *Information Technology*, *4*(8), 327-329.

Zhang, D., Jiang, Q., & Li, X. (2007). Application of neural networks in financial data mining. *International Journal of Computer and Information Engineering*, *1*(1), 225-228.