

# Performance Analysis of Stock Market Prediction Techniques

Roshani Gandhe<sup>1</sup> Prof. Manoj S. Chaudhari<sup>2</sup>

<sup>1</sup>M.Tech. Student <sup>2</sup>Professor & HOD

<sup>1,2</sup>Department of Computer Science & Engineering

<sup>1,2</sup>Priyadarshini Bhagwati College of Engineering, Nagpur, India

*Abstract*— Stock market values keeps on changing day by day, so it is very difficult to predict the future value of the market. Although there are various techniques implemented for the prediction of stock market values, but the predicted values are not very accurate and error rate is more. Hence an efficient technique is implemented for the prediction of the stock market values using hybrid combinatorial method of clustering and classification. The dataset is taken from shanghai stock exchange market and is first clustered using K-means clustering algorithm and these clustered values are classified using horizontal partition decision tree. Stock prices prediction is interesting and challenging research topic. Developed countries' economies are measured according to their power economy. Currently, stock markets are considered to be an illustrious trading field because in many cases it gives easy profits with low risk rate of return. Stock market with its huge and dynamic information sources is considered as a suitable environment for data mining and business researchers. In this paper, we applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock prices for a sample of six major companies listed on the Jordanian stock exchange to assist investors, management, decision makers, and users in making correct and informed investments decisions. According to the results, the kNN algorithm is robust with small error ratio; consequently the results were rational and also reasonable. In addition, depending on the actual stock prices data; the prediction results were close and almost parallel to actual stock prices.

**Key words:** Stock Price Prediction, Listed Companies, Data Mining, K-Nearest Neighbor, Non-Linear Regression

## I. INTRODUCTION

Stock price forecasting is an important task for investment and financial decision making. Researches and practitioners have given it considerable amount of attention. Stock market is most volatile and highly risky investment. Many approaches have been used for forecasting stock price such as traditional and fundamental methods. Forecasting stock price or financial markets has been one of the biggest challenges to the AI community. A variety of fundamental, statistical, and technical indicators have been offered and used with changeable results but these previous methods have limits and not completely capable to provide accuracy. Data mining and computational intelligence techniques for resolving the problems of stock future price forecasting have become rapidly growing alternative methods for achieving considerable degree of accuracy. The purpose of forecasting research has been largely beyond the capability of traditional AI research which has mainly focused on developing intelligent systems that are supposed to emulate human intelligence [1].

From the beginning of time it has been man's common goal to make his life easier. The prevailing notion in society is that wealth brings comfort and luxury, so it is

not surprising that there has been so much work done on ways to predict the markets. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. Analyzing the company's operations and the market in which the company is operating can do this. Consequently, the stock price can be predicted reasonably well. Most people believe that fundamental analysis is a good method only on a long-term basis. However, for short- and medium term speculations, fundamental analysis is generally not suitable. Past approaches to this problem first applied an artificial neural network directly to historical stock data using linear time series modelling. This produced results which over fitted the training data and therefore rendered them unusable in real trading. Additionally to omitting any preprocessing, the neural networks employed only restricted two layers, an output and an input layer. These linear techniques are now known to be provably insufficient for any nonlinear phenomenon including stock price movement. Some other research used the techniques of technical analysis, in which trading rules were developed based on the historical data of stock trading price and volume. Technical analysis as illustrated in and refers to the various methods that aim to predict future price movements using past stock prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data. Thus, it is assumed that price trends and patterns exist that can be identified and utilized for profit. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid.

Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area. Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis. Data classification can be done in many different methods; one of those methods is the classification by using Decision Tree. It is a graphical representation of all possible outcomes and the paths by which they may be reached. Decision trees and artificial neural networks can be trained by using an appropriate learning algorithm. Once these patterns have been discovered, future prices can be predicted. Today, the grand challenge of using a database is to generate useful rules from raw data in a database for users to make decisions, and these rules may be hidden deeply in the raw data of the database. Traditionally, the method of turning data into knowledge relies on manual analysis; this is becoming impractical in many domains as data volumes grow exponentially. The problem with predicting stock prices is that the volume of data is too large and huge. The

classification approach on the historical data available to try to help the investors to build their decision on whether to buy or sell that stock in order to achieve profit.

The main objective of this paper is to analyze the historical data available on stocks using decision tree technique as one of the classification methods of data mining in order to help investors to know when to buy new stocks or to sell their stocks. Analysing stock price data over several years may involve a few hundreds or thousands of records, but these must be selected from millions. The data that will be used to build the decision tree will be the historical prices of Stock Exchange over two years of time.

## II. LITERATURE REVIEW

Chenoweth Tim, Obradovic, and Sauchi [3] relied on a single technical indicator called the average direction index (ADX), which identifies and quantifies trends by averaging the fraction of today's range above or below the earlier day's assortment. The ADX is achieved through a feature selection component and used as input into two separate neural networks (Up and Down) whose results were then polled and applied to a rule base to predict the final market movement. Exclusive of knowing the accurate predictive accurateness, it is complicated to quantitatively judge against the system, which unavoidably comprises rules for trading which may be the actual cause of the monetary gain achieved by the system rather than predictive accuracy. Roman and Akhtar [4] investigate that if back propagation and recurrent neural networks can be effectively used in designing portfolios across many international stock markets after the trends in these markets for several calendar years are known. Stock prices fluctuate daily resulting in a nonlinear pattern of data.

Shaun and Ruey [5] proposes a stock market forecasting system based on artificial neural network. They train the system with 500 composite indexes of past twenty years. The system produces the forecast and adjust itself by comparing its forecasts with the actual indexes. They also develop a transfer function model to forecast based on the indexes and the forecasts by the artificial neural networks. Huang, Nakamori, and Shou [6] propose SVM based stock market prediction technique. They compare its performance with Quadratic Discriminant Analysis, Linear Discriminant Analysis and Elman Back-Propagation Neural Network. After comparing them they also propose a combining model by integrating SVM with other classification method.

Mahfoud, Sam, and Mani [7] use genetic algorithms to predict stock prices. Genetic algorithms are encouraged by evolutionary biology and the concept of survival of the fittest. A bulky population of probable algorithmic representations for stock prediction is first produced. Then, each member is implemented and assessed, keeping the algorithms that produce the best results and mixing their properties amongst other high scoring algorithms to obtain a new generation of algorithms in a Darwinian fashion. The process is repetitive until the error has been reduced to an acceptable level.

Naeini, Hamidreza, and Homa [8] propose a neural network stock market prediction technique. They use two kinds of neural networks a feed forward multilayer perception (MLP) and an Elman recurrent network. They found that MLP neural network is much better in predicting

stock value changes than Elman recurrent network and linear regression method.

Wang, Long, and Chan [9] propose stock prediction technique based on rule discovery. It uses a two-layer bias decision tree. The technique used in this study differs from other studies in two aspects. First, this study modified the decision model into the bias decision model to reduce the classification error. Second, this study uses the two-layer bias decision tree to improve purchasing accuracy. This technique improve purchasing accuracy, investment returns and also have the advantages of fast learning speed robustness, stability, and generality.

### A. Disadvantages of the Existing Systems

- 1) Stocks are volatile investments. The price of a single stock can vary quite widely from day to day, and the factors that cause these price fluctuations are beyond the control of the investor.
- 2) Buying a widely diversified basket of stocks can be difficult for all but the wealthiest investor. Small investors are better off buying a quality stock mutual fund. Mutual funds pool the investments of many different people in order to buy a diversified set of stocks. This diversified approach helps to reduce the risk inherent in the stock market.
- 3) As investors near retirement, the amount of stocks in the portfolio should be reduced. Investors who are close to retirement age can no longer afford to take chances with their money, and that means moving a significant portion of their retirement funds to safer and more stable investments.
- 4) Buying and selling stocks costs money in the form of brokerage commissions, and many brokerage firms charge account maintenance fees as well. It is important to look for low cost alternatives when buying and selling stocks.
- 5) Extensive attention has been dedicated to the field of artificial intelligence and data mining techniques. Some models have been proposed and implemented using the above mentioned techniques.

## III. OBJECTIVE

It may seem overly presumptuous to believe that historical price fluctuations alone can be used to predict the direction of future prices. It may thus seem natural to include some fundamental analysis knowledge in the feature generation process. However, due to the inherent limitations in time and the added complexity of including a second analysis technique, this has not been a priority. We have instead placed focus on creating a model that can be easily extended with new analysis techniques, not necessarily from technical analysis, by using two separate reasoning layers and using an agent-oriented approach for the domain knowledge. Our goal is not to justify, prove or disprove technical analysis; by focusing strictly on technical indicators we are presented with an opportunity to evaluate the utility of selected methods in this form of stock analysis.

- The main reason of building the model is to try to help the investors in the stock market to decide the best timing for buying or selling stocks based on the knowledge extracted from the historical prices of such stocks.

- The decision taken will be based on one of the data mining techniques.
- To study the current stock market trend and collect trend data.
- To build prediction model for the companies listed under index of NSE using multilayer perception (MLP) Neural Network technique.
- To compare the model with real data for its accuracy.
- These continuous trends can be analyzed and manipulated to predict the changes in the stock for making investment strategy.

#### A. Proposed System

Accordingly, changes in the stock price reflect release of new information, changes in the market generally, or random movements around the value that reflects the existing information set.

- There are two prices that are critical for any investor to know: the current price of the investment he or she owns, or plans to own, and its future selling price. Despite this, investors are constantly reviewing past pricing history and using it to influence their future investment decisions.
- Some investors won't buy a stock or index that has risen too sharply, because they assume that it's due for a correction, while other investors avoid a falling stock, because they fear that it will continue to deteriorate. Here that our assumption is lots of prediction algorithms are used. So here we used base of clustering and classification of algorithm to implement stock market prediction.
- The stock market is one of the most important ways for investors to raise money, along with debt markets which are generally more imposing but do not trade publicly.
- History has shown that the price of stocks and other assets is an important part of the dynamics of economic activity, and can influence or be an indicator of social mood. An economy where the stock market is on the rise is considered to be an up-and-coming economy. In fact, the stock market is often considered the primary indicator of a country's economic strength and development.
- Rising share prices, for instance, tend to be associated with increased business investment and vice versa. Share prices also affect the wealth of households and their consumption. Therefore the investors to keep an eye on the control and behavior of the stock market and, in general, on the smooth operation of financial system functions.
- The smooth functioning of all these activities facilitates economic growth in that lower costs and enterprise risks promote the production of goods and services as well as possibly employment. In this way the financial system is assumed to contribute to increased prosperity, although some controversy exists as to whether the optimal financial system is bank-based or market-based.

## IV. RESEARCH METHODOLOGY AND ANALYSIS

### A. Implementation

#### 1) HTML Parser

Extraction encompasses all the information retrieval programs that are not meant to preserve the source page. This covers uses like: text extraction, for use as input for text search engine databases for example link extraction, for crawling through web pages or harvesting email addresses screen scraping, for programmatic data input from web pages resource extraction, collecting images or sound a browser front end, the preliminary stage of page display link checking, ensuring links are valid site monitoring, checking for page differences beyond simplistic differs. There are several facilities in the HTML Parser codebase to help with extraction, including filters, visitors and JavaBeans. Transformation includes all processing where the input and the output are HTML pages.

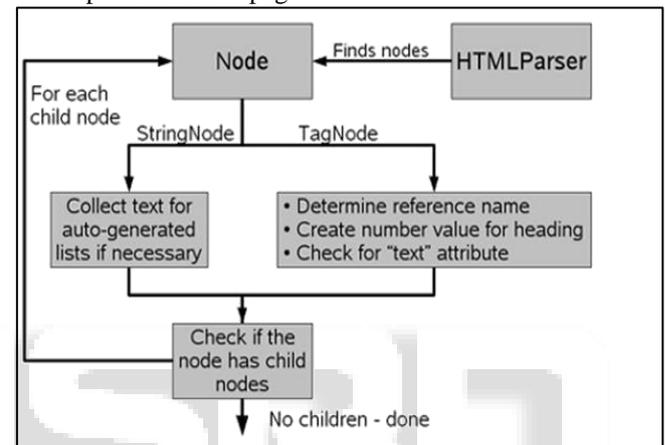


Fig. 1: HTML Parser

Some examples are: URL rewriting, modifying some or all links on a page site capture, moving content from the web to local disk censorship, removing offending words and phrases from pages HTML clean-up, correcting erroneous pages ad removal, excising URLs referencing advertising conversion to XML, moving existing web pages to XML During or after reading in a page, operations on the nodes can accomplish many transformation tasks "in place", which can then be output with the Html () method. Depending on the purpose of your application, you will probably want to look into node decorators, visitors, or custom tags in conjunction with the Prototypical Node Factory.

### B. Data Mining

Data mining methodology is designed to ensure that the data mining effort leads to a stable model that successfully addresses the problem it is designed to solve. Various data mining methodologies have been proposed to serve as blueprints for how to organize the process of gathering data, analysing data, disseminating results, implementing results, and monitoring improvements. This model consists of the following six steps:

- Understanding the reason and objective of mining the stock prices.
- Understanding the collected data and how it is structured.
- Preparing the data that is used in the classification model.

- Selecting the technique to build the model.
- Evaluating the model by using one of the well-known evaluation methods.
- Deploying the model in the stock market to predict the best action to be taken, either selling or buying the stocks.
- Understanding the reason and objective of building the model.

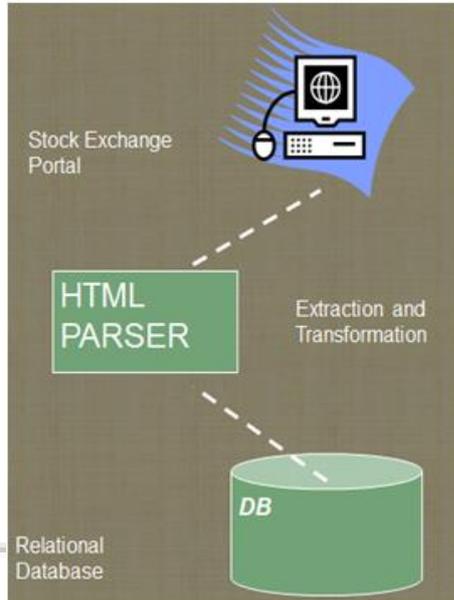


Fig. 2: System Architecture

A data mining system can be classified according to the kinds of databases mined. Database systems themselves can be classified according to different criteria (such as data models, or the types of data or applications involved), each of which may require its own data mining technique. Data mining systems can therefore be classified accordingly. For instance, if classifying according to data models, we may have a relational, transactional, object-oriented, object-relational, or data warehouse mining system. If classifying according to the special types of data handled, we may have a spatial, time-series, text, or multimedia data mining system, or a World-Wide Web mining system. Other system types include heterogeneous data mining systems, and legacy data mining systems. Data mining systems can also be categorized according to the underlying data mining techniques employed. These techniques can be described according to the degree of user interaction involved (e.g., autonomous systems, interactive exploratory systems, query-driven systems), or the methods of data analysis employed (e.g., database-oriented or data warehouse-oriented techniques, machine learning, statistics, visualization, pattern recognition, neural networks, and so on). A sophisticated data mining system will often adopt multiple data mining techniques or work out an effective, integrated technique which combines the merits of a few individual approaches.

### C. K-Means Method

K-means is one of the simplest unsupervised learning algorithms that solve the clustering problem. The algorithm executes in simple way by classifying a given data set through a certain number of clusters (say k clusters) fixed Apriori. The basic concept is to define k centers; one for

each cluster. So, it is better way to place them as much as possible far away from each other. In next step it takes each point belonging to a given data set and associates it to the nearest center. When no point remains, the first step is completed and an early grouping is done.

Here we need to re-calculate k new centroids as barycentre of the clusters obtained from the step.

- Select 'c' cluster centers randomly.
- Calculate the distance b/w each data point and cluster centers.
- Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
- Again calculate the new cluster center.
- Again calculate the distance between each data point and new obtained cluster centers.
- If there is no reassignment of data points then stop, otherwise repeat from step 3.

### D. Data Base Server

The term database server may refer to both hardware and software used to run a database, according to the context. As software, a database server is the back-end portion of a database application, following the traditional client-server model. This back-end portion is sometimes called the instance. It may also refer to the physical computer used to host the database. When mentioned in this context, the database server is typically a dedicated higher-end computer that hosts the database. A database server is a computer program that provides database services to other computer programs or computers, as defined by the client-server model. The term may also refer to a computer dedicated to running such a program. Database management systems frequently provide database server functionality, and some DBMSs (e.g., MySQL) rely exclusively on the client-server model for database access.

Such a server is accessed either through a "front end" running on the user's computer which displays requested data or the "back end" which runs on the server and handles tasks such as data analysis and storage. In a master-slave model, database master servers are central and primary locations of data while database slave servers are synchronized backups of the master acting as proxies. Most of the Database server works with the base of Query language. Each Database understands its query language and converts it to Server readable form and executes it to retrieve the results.

Some examples of proprietary database servers are Oracle, DB2, Informix, and Microsoft SQL Server. Examples of GNU General Public License database servers are Ingres and MySQL. Every server uses its own query logic and structure. The SQL query language is more or less the same in all relational database servers. DB-Engines list over 200 DBMSs in its ranking. In the client-server computing model, there is a dedicated host to run and serve up the resources, typically one or more software applications. There are also several clients who can connect to the server and use the resources offered and hosted by this server.

## V. EXPERIMENTAL RESULT

Predicting the results of the stock market should not be synonymous to that to predicting weather conditions. So, how does the software help predict results of the stock market compare to the human element.

- The stock market witnessed a major transformation and structural change during the period.
- The stock market fluctuates by the external environment.
- Stock market is all about future prediction.
- Stock is very sensitive market.
- It is based on "high risk and high return".
- Buying and selling of shares through the internet is so easy.
- There are 57% people are doing trading in stock market.
- There are 59% people are doing trading in equity and others are in commodity, mutual fund.

## VI. CONCLUSIONS

The proposed technique implemented here to study to presents a proposal to use the decision tree classifier on the historical prices of the stocks to create decision rules that give buy or sell recommendations in the stock market. Such proposed model can be a helpful tool for the investors to take the right decision regarding their stocks based on the analysis of the historical prices of stocks in order to extract any predictive information from that historical data. The results for the proposed model were not perfect because many factors including but not limited to political events, general economic conditions, and investors' expectations influence stock market.

Finally, reconsidering the factors affecting the behavior of the stock markets, such as trading volume, news and financial reports which might impact stock price can be another rich field for future studying. In Data Mining to predict stock market here we have created NLP based module & statistical parameter based module which results the sentence polarity & behavior compared to past year data. By using this technique we get accurate & reliable prediction result which give consumer better solution for where to invest their valuable money. These modules evaluate the news sentences based on grammatical analysis and with the help of historical data also.

## REFERENCES

- [1] Wang, Y.F., (2003) "Mining stock price using fuzzy rough set system", *Expert Systems with Applications*, 24, pp. 13-23.
- [2] Wu, M.C., Lin, S.Y., and Lin, C.H., (2006) "An effective application of decision tree to stock trading", *Expert Systems with Applications*, 31, pp.270-274.
- [3] Chenoweth, Tim, Zoran Obradovic, and Sauchi Stephen Lee. "Embedding technical analysis into neural network based trading systems." *Applied Artificial Intelligence* 10.6: 523-542, 1996.
- [4] Roman, Jovina, and Akhtar Jameel. "Back propagation and recurrent neural networks in financial analysis of multiple stock market returns." *System Sciences*, 1996, Proceedings of the TwentyNinth Hawaii International Conference on., Vol. 2. IEEE, 1996.
- [5] Wu, S. and Lu, R. P. "Combining artificial neural networks and statistics for stock market forecasting", *Proceedings of the 1993 ACM conference on Computer*, pp. 257 – 264, 1993.
- [6] Huang, Wei, Yoshiteru Nakamori, and Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine." *Computers & Operations Research* 32, no. 10: 2513-2522, 2005.
- [7] Mahfoud, Sam, and Ganesh Mani. "Financial forecasting using genetic algorithms." *Applied Artificial Intelligence* 10, no. 6, 543-566, 1996.
- [8] Naeini, Mahdi Pakdaman, Hamidreza Taremian, and Homa Baradaran Hashemi. "Stock market value prediction using neural networks." *Computer Information Systems and Industrial Management Applications (CISIM)*, 2010 International Conference on. IEEE, 2010.
- [9] Wang, Jar-Long, and Shu-Hui Chan. "Stock market trading rule discovery using two-layer bias decision tree." *Expert Systems with Applications* 30.4, 605-611, 2006.
- [10] AlDebie, M., Walker, M. (1999). "Fundamental information analysis: An extension and UK evidence", *Journal of Accounting Research*, 31(3), pp. 261–280.
- [11] Lev, B., Thiagarajan, R. (1993). "Fundamental information analysis", *Journal of Accounting Research* 31(2), 190–215.
- [12] Tsang, P.M., Kwok, P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K., and Wong, T.L. (2007) "Design and implementation of NN5 for Hong Kong stock price forecasting", *Engineering Applications of Artificial Intelligence*, 20, pp. 453-461.
- [13] Ritchie, J.C., (1996) *Fundamental Analysis: a Back-to-the-Basics Investment Guide to Selecting Quality Stocks*. Irwin Professional Publishing.
- [14] Murphy, J.J., (1999) *Technical Analysis of the Financial Markets: a Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance.
- [15] Wang, Y.F., (2002) "Predicting stock price using fuzzy grey prediction system", *Expert Systems with Applications*, 22, pp. 33-39.
- [16] Han, J., Kamber, M., Jian P. (2011). "Data Mining Concepts and Techniques". San Francisco, CA: Morgan Kaufmann Publishers.
- [17] Enke, D., Thawornwong, S. (2005) "The use of data mining and neural networks for forecasting stock market returns", *Expert Systems with Applications*, 29, pp. 927-940.
- [18] Wang, J.L., Chan, S.H. (2006) "Stock market trading rule discovery using two-layer bias decision tree", *Expert Systems with Applications*, 30(4), pp.605-611.
- [19] Lin, C. H. (2004) Profitability of a filter trading rule on the Taiwan stock exchange market. Master thesis, Department of Industrial Engineering and Management, National Chiao Tung University.
- [20] Cao, Q., Leggio, K.B., and Schniederjans, M.J., (2005) "A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market", *Computers & Operations Research*, 32, pp. 2499-2512.

- [21] Fama, E.F., French, K.R., (1993) "Common risk factors in the returns on stocks and bonds", *The Journal of Finance*, 33, pp. 3-56.
- [22] Fama, E.F., French, K.R., (1992) "The cross-section of expected stock returns", *The Journal of Finance*, 47, pp. 427-465.
- [23] Al-Haddad W. Alzurqan S. and Al\_Sufy S, The Effect of Corporate Governance on the Performance of Jordanian Industrial Companies: An empirical study on Amman Stock Exchange. *International Journal of Humanities and Social Science*, Vol. 1 No. 4; April 2011.
- [24] Hajizadeh E., Ardakani H., and Shahrabi J., Application of data mining techniques in stock markets: A survey, *Journal of Economics and International Finance* Vol. 2(7), pp. 109-118, July 2010.
- [25] Soni S., Applications of ANNs in Stock Market Prediction: A Survey, *International Journal of Computer Science & Engineering Technology (IJCSET)*, pp 71-83, Vol. 2 No. 3, 2011.
- [26] Hazem M. El-Bakry, and Wael A. Awad, Fast Forecasting of Stock Market Prices by using New High Speed Time Delay Neural Networks, *International Journal of Computer and Information Engineering* 4:2 2010. Pp 138-144.
- [27] Witten I. Frank E., and Hall M. (2011), "Data Mining: Practical Machine Learning Tools and Techniques", 3rd Edition, Morgan Kaufmann Publishers.
- [28] Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., and Wirth R., (2000). "CRISP-DM 1.0: Step-by-step data mining guide".

