Prediction of Market Capital for Trading Firms through Data Mining Techniques

Aditya Nawani
Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi, India

Himanshu Gupta
Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi, India

Narina Thakur
Department of Computer Science, Bharati Vidyapeeth's College of Engineering, New Delhi, India

ABSTRACT

The ultimate goal of data mining is prediction- and predictive data mining is the most common type of data mining and one that has most direct business applications. This paper discusses how one can apply data mining to design a market capital prediction system for trading firms[1]. The dataset is normalized and trained. The paper delves into the field of neural networks and shows how it can be utilized, in combination with the Graphical user Interface of MATLAB, GUIDE, to make accurate predictions. When implemented, the trained system can be used to forecast the market capital for a particular combination of input parameters. The accuracy of this method demonstrates its utility as a predictive tool.

General Terms

Classification, Prediction, Neural Network

Keywords

Data Mining, MATLAB, ANN, Prediction, Neural network

1. INTRODUCTION

1.1 Data mining

Data mining is an analytic process designed to explore data (usually large amount of data –typically business or market related –also known as "big data"), seeking steady and unwavering relationships between the various parameters[2]. Once the relationships are determined, the next step is to validate them using a test data. The process of data mining consists of:

1) Pattern finding

2) Pattern verification/validation

3) Application of the learned set for further predictions

1.2 Neural networks

An **Artificial Neural Network (ANN)**, often just called a neural network, is a set of interconnected links that have weights associated with them[3]. The concept of ANN was derived from biological neural networks, and is the cynosure of this paper.

Neural networks open up a new foray into the field of making efficient and usable predictions in order to optimize profits. Artificial Neural Networks are being used in numerous areas, as it is an irrefutably effective tool that aids the scientific community in forecasting about probable outcomes[4].

Any ANN can be thought of as a set of interconnected units broadly categorized into three layers. These three layers are the input layer, the hidden layer and the output layer. Inputs are fed into the input layer, and its weighted outputs are passed onto the hidden layer. The neurons in the hidden layer (hidden neurons) are essentially concealed from view. Using additional levels of hidden neurons provides increased flexibility and more accurate processing. However, the flexibility comes at the cost of extra complexity in the training algorithm. Having more hidden neurons than necessary is wasteful, as a less number of neurons would serve our purpose just fine. On the other hand, having less hidden neurons than required would cause reduced robustness of the system, and defeat its very purpose.

An illustration of a neural network is shown in Figure 1.

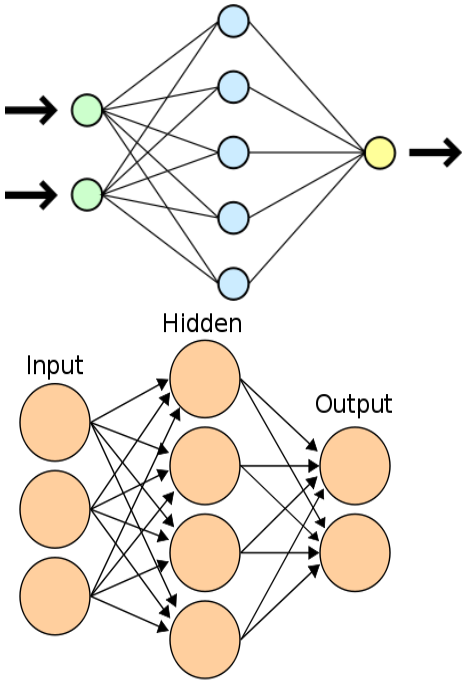

Fig 1: A Neural Network system

1.3 Neural network applications

- **Character Recognition** – Neural networks can be used to identify handwritten characters. This property gains

significance by the day as we witness a surge in the sales of handheld devices.

- **Image Compression** – People deal with vast amounts of data nowadays, and compression of data is invariably used to send or receive images, text et cetera. Neural networks play a role in image compression.
- **Stock Market Prediction** – The functioning of the stock market is quite convoluted, and is subject to a plethora of determining factors. Since neural networks can process a lot of information in a short span of time, they can be used to forecast stock values.
- **Travelling Salesman's Problem** - Interestingly enough, neural networks can solve the travelling salesman problem, but only to a certain degree of approximation.
- **Medicine, Electronic Nose, Security, and Loan Applications** - These are some applications that are in their proof-of-concept stage, involving a neural network that will decide whether or not to grant a loan.
- **Miscellaneous Applications** - There are some other very interesting applications of neural networks[6][7].

2. IMPLEMENTATION

- The dataset was obtained online and comprised a large collection of attributes, chiefly Industry Group, Country, Sub Group, Broad Group and Market Capital. It contained information about companies from countries all around the world. The Industry Group constituted the major areas that global trading firms deal with. It covered many areas from electronics to apparel. The country, sub group and broad group parameters represented the country the firm belongs to and the major and minor groups the firm falls into respectively.
- A numeric value was assigned to each of the input entries. This way, all industries belonging to each and every category and country had a unique numeric identity.
- The text file was converted to a .csv (comma separated value) file and fed as input to the system. Since non-normalized values of the target file would always yield erroneous results, the values had to be normalized first. This is accomplished using Microsoft Excel. Excel has some very useful tools and functions. A few of them are the STDDEV, STANDARDIZE and AVERAGE functions. The values got normalized, that is, they were uniformly distributed about a middle point. Half of the values lay to the left of the middle point, and were negative. The other half lay to the other side, and consisted of positive numbers. This data was again converted to a .csv file. The result was an input file and a target file that could be used for the purpose of making accurate predictions. Henceforth, the next steps would be to train the data, and create a GUI for the system.
- The neural network tool is a very helpful tool that permits us to train a dataset, so that the network can intelligently predict future values. Four neurons were selected for the first layer and one neuron was selected for the second layer, and the network type was chosen as Feed Forward Backprop. Then the network could be simulated.
- To facilitate easy implementation, the network file was loaded into a function, and then this function was called within the GUI. The GUI itself was created in MATLAB, using GUIDE (GUI development environment).

  Figure 2 and Figure 3 show respectively the input and normalized target dataset being fed into the neural network.

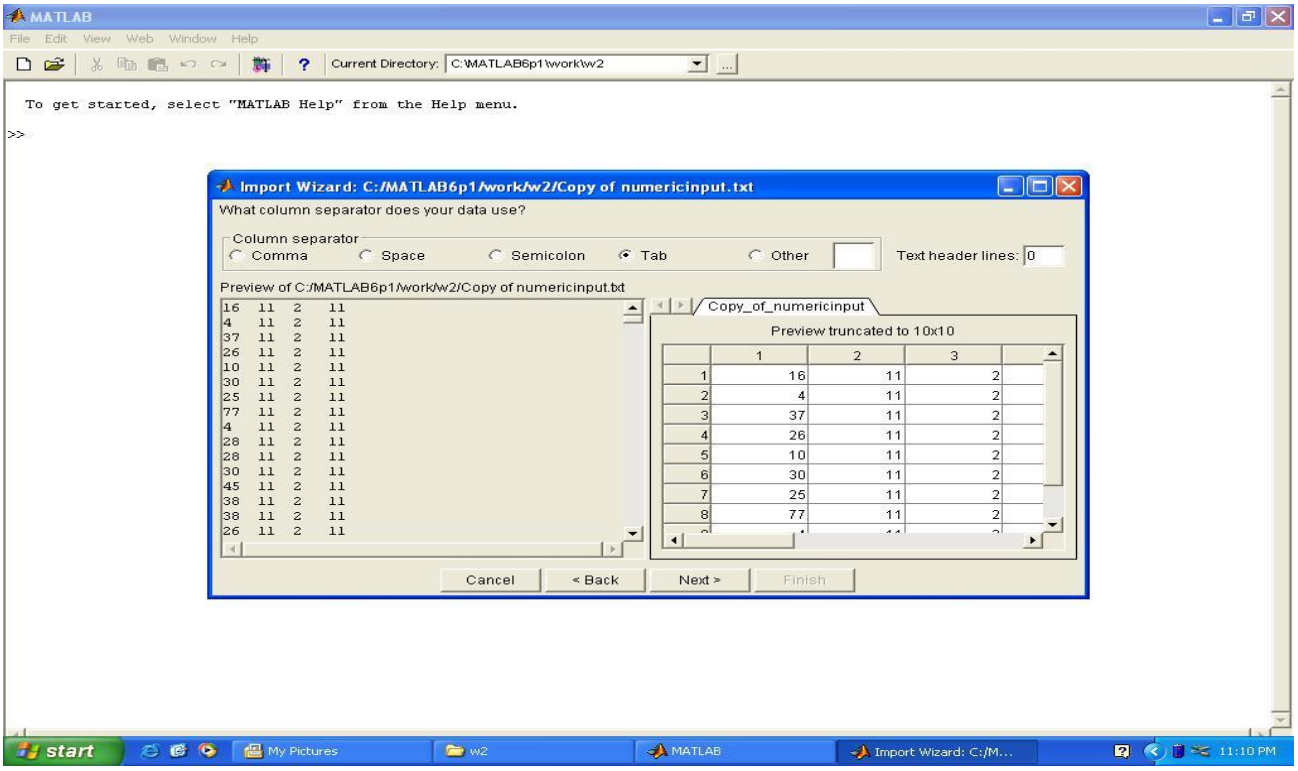

Fig 2: The input dataset

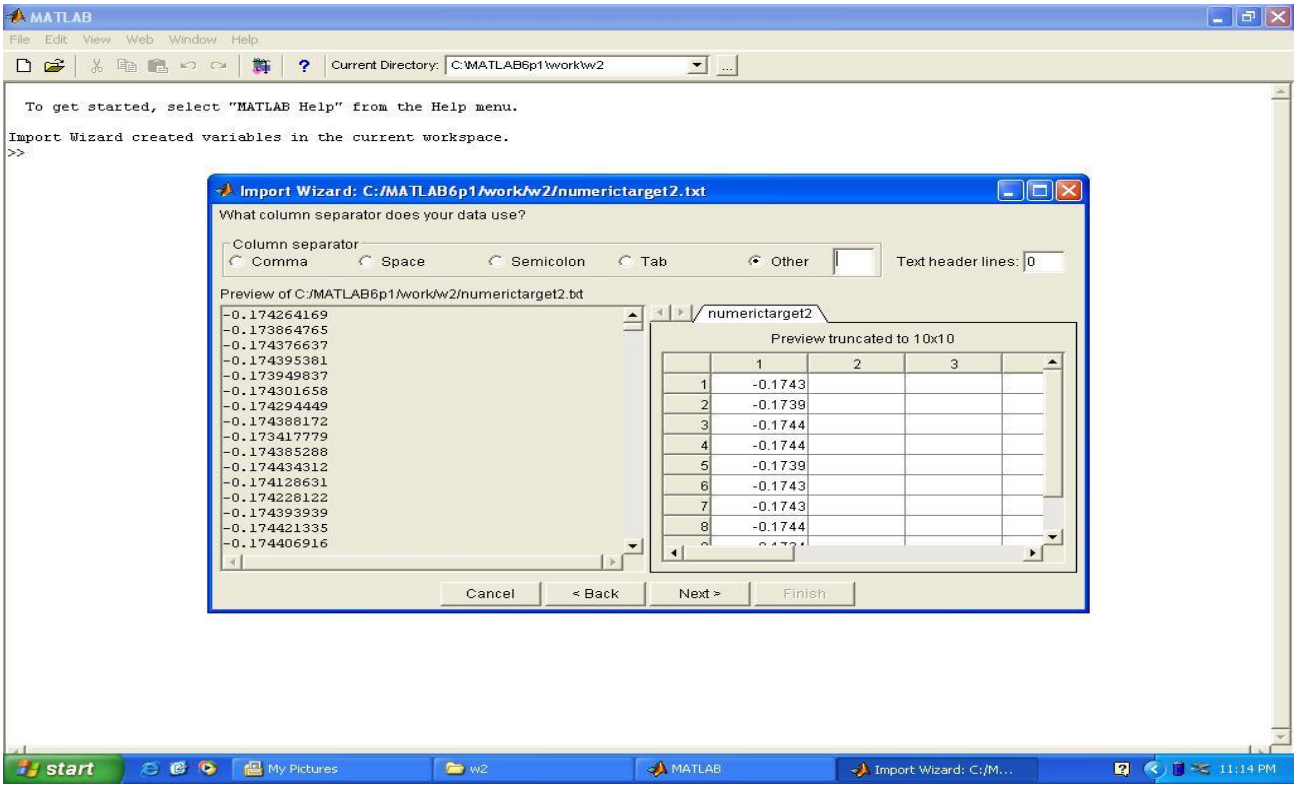

Fig 3: The preprocessed and normalized data

Figure 4 is a snapshot of the neural network created using the neural network tool, whereas Figure 5 shows the process of specifying the input and the target to be used for the system.

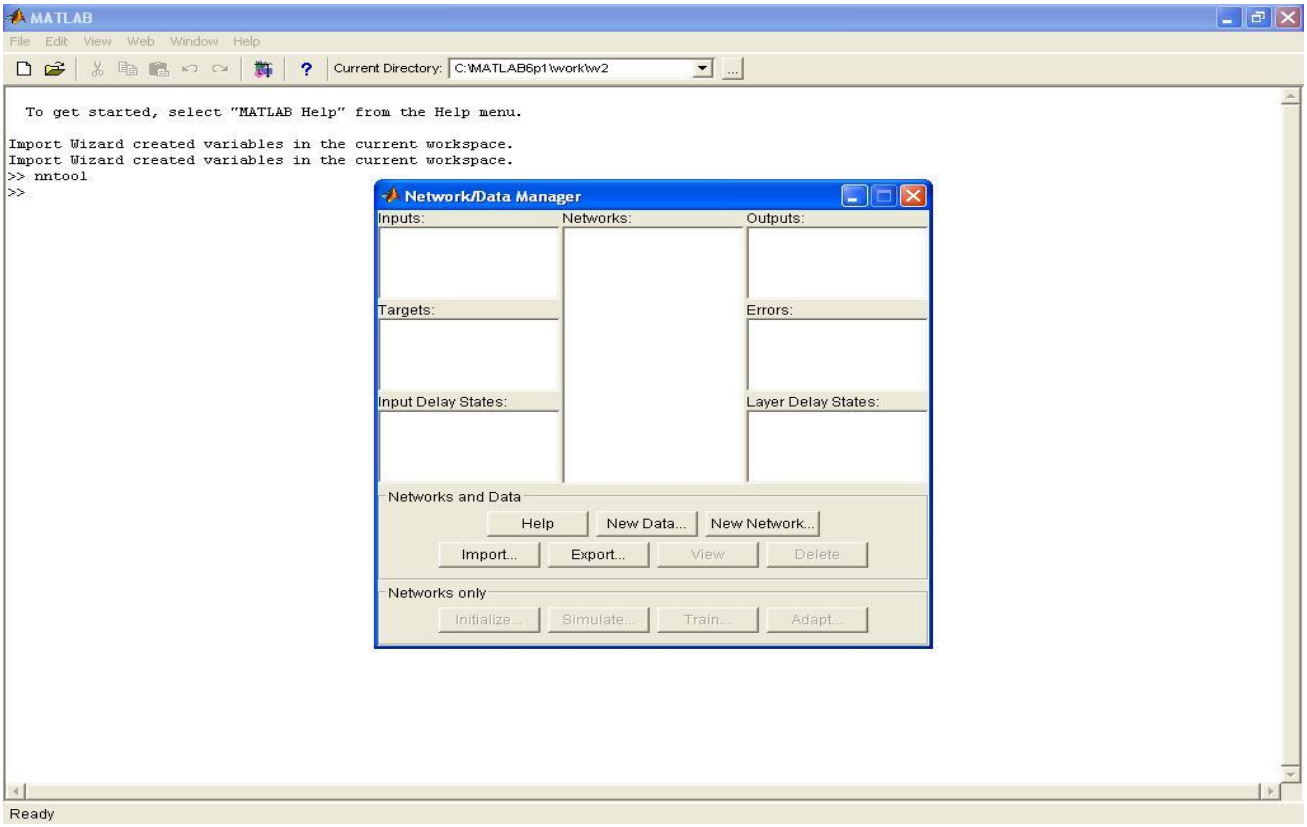

Fig 4: Creating a network

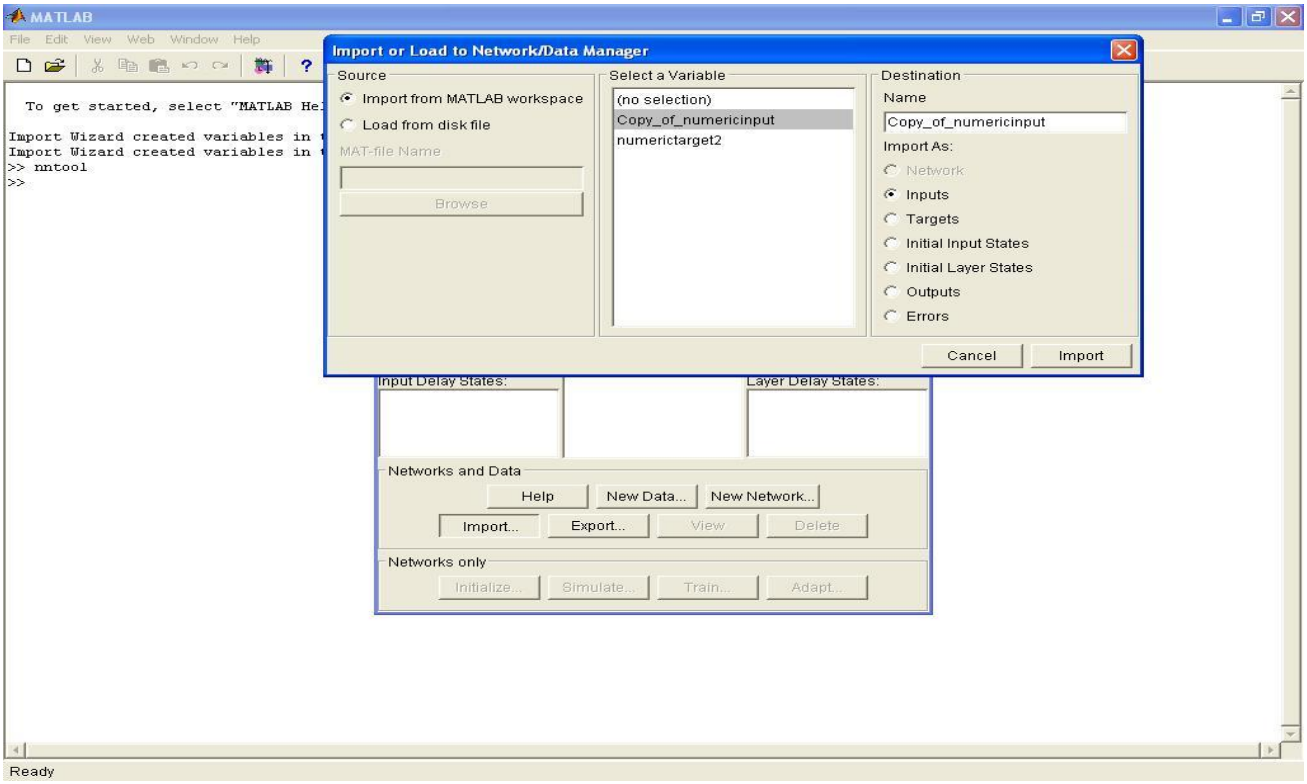

Fig 5: Specifying the input and the target

Figure 6 shows the process of selecting the number of neurons to be used for each layer of the system. A pictorial view of the artificial neural network is given in Figure 7.

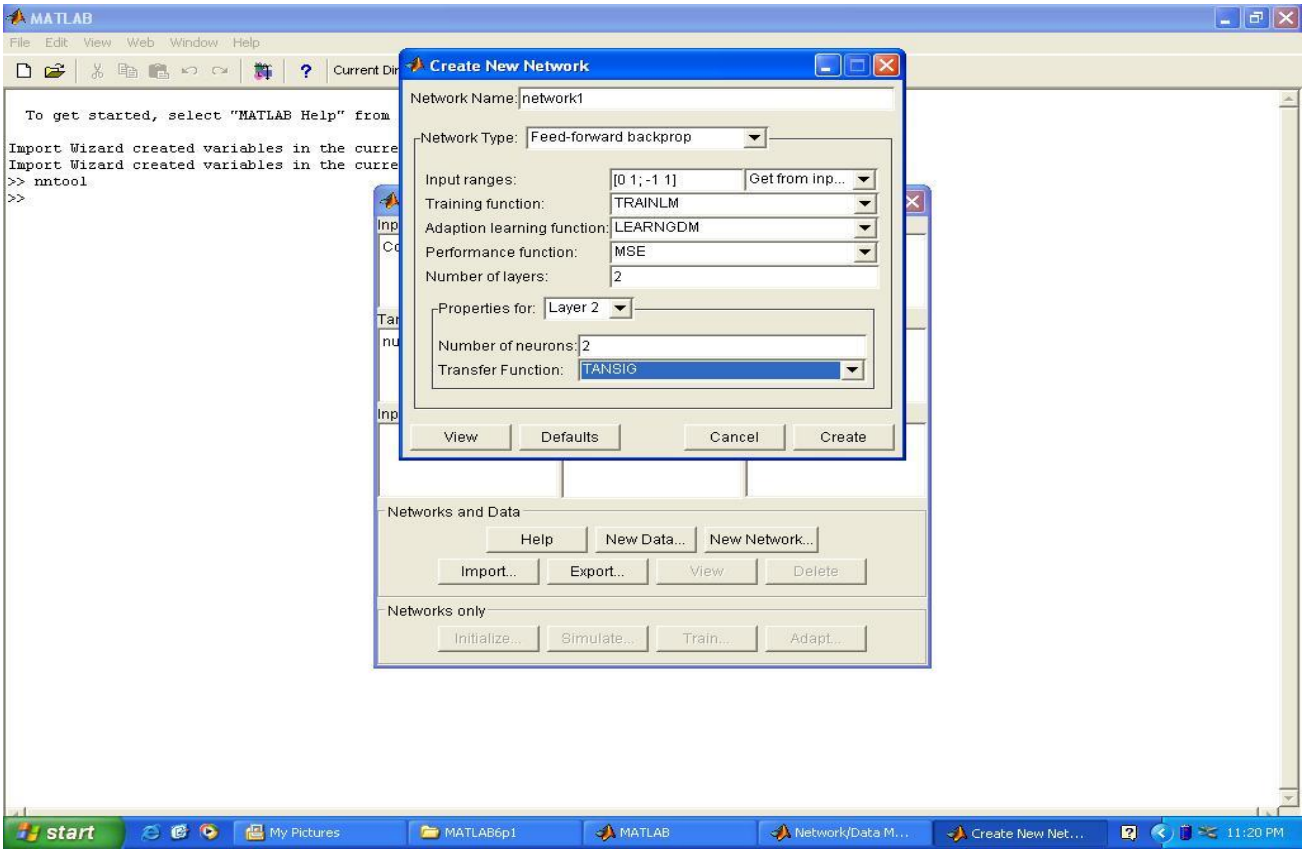

Fig 6: Specifying the number of neurons and transfer function

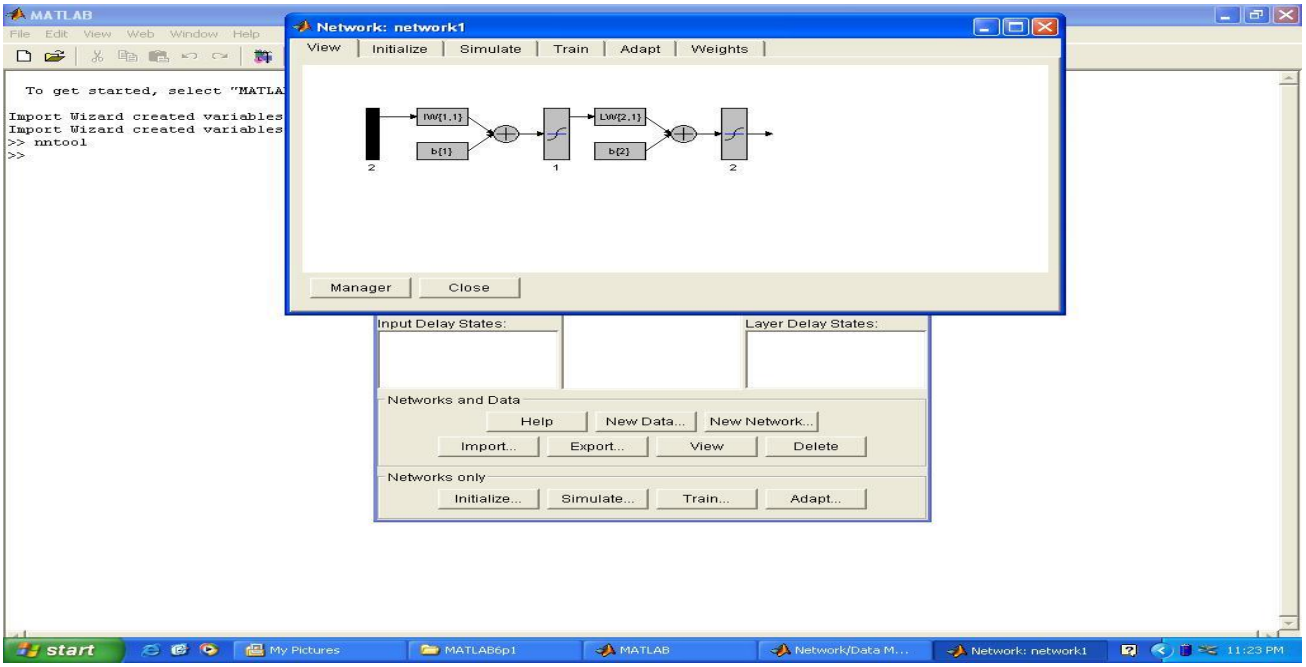

Fig 7: A view of the created network

The GUI of the Trading Firms Market Capital Prediction System is shown by Figure 8.

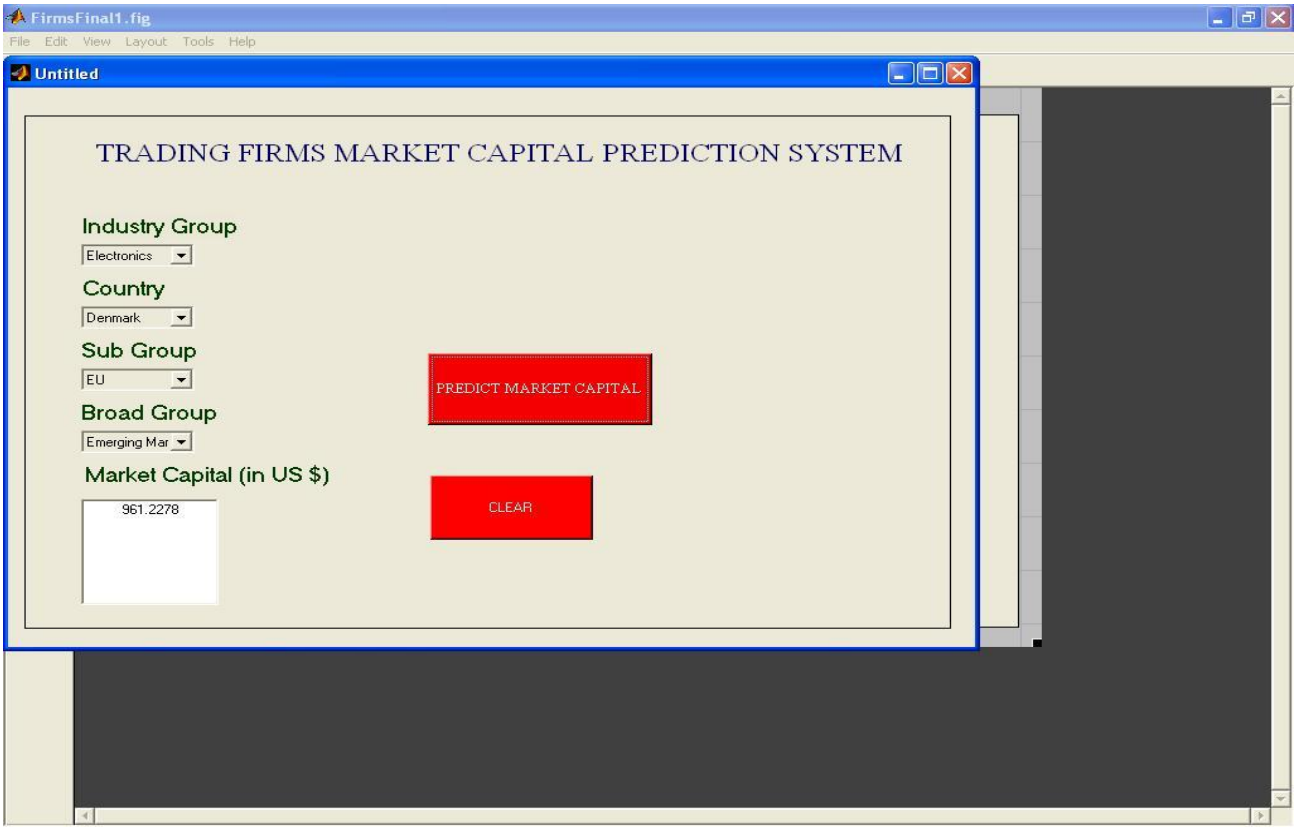

Fig 8: GUI of the Trading Firms market Capital Prediction System

We also constructed a graph of the predicted market capital for any given set of input parameters, and an instance of the linear plot is shown by Figure 9.

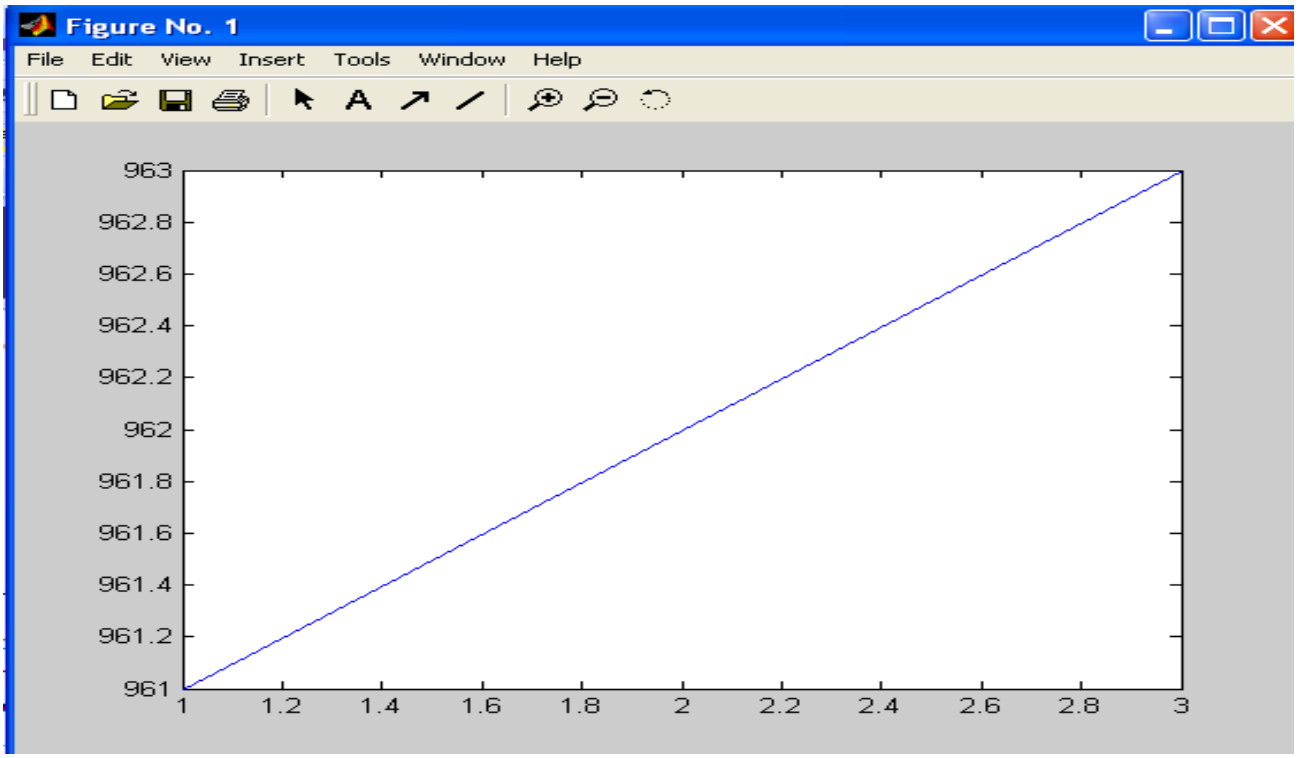

Fig 9: A linear plot of the predicted market capital

3. RESULTS AND ANALYSIS

At first glance, the results obtained were found to be comparable to the output expected. Further investigation shows that the outcome is indeed almost similar to that anticipated. More comparison can be done with statistical techniques like moving average, regression and by simple analysis of trading firms and market trends.

4. CONCLUSION AND FUTURE SCOPE

In the world of finance and global commerce, prediction of the returns of investing in a particular firm is a matter of the utmost importance[8][9]. For long, artificial neural networks have been used in the field of prediction. Sometimes, it has been found that artificial neural networks possess drawbacks when learning data patterns. They have also been known to demonstrate inconsistent and unpredictable behavior if the data used is too massive or complex. However, the overall percentage of errors or deviations from the result expected being low, it can be safely concluded that artificial neural networks have a vast future scope in the domain of economics and prediction. Any smart investor would definitely like to refer to such an intelligent system to maximize his profits. Hence, we could be seeing a greater involvement of artificial neural networks in foretelling global market trends in the future.

5. REFERENCES

[1] J. Wang, 2008. Multivariate Statistical Analysis, Science publishing Company.

[2] P. Tan, M. Steinbach and V. Kumar, 2006. Introduction to Data Mining, The People's Posts and Telecommunications Press.

[3] Kartalopoulos S.V., 2000. Understanding Neural Networks and Fuzzy Logic, Prentice-Hall.

[4] Jang, J.-S. R., 1991. "Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm," Proc. of the Ninth National Conf. on Artificial Intelligence (AAAI-91), pp. 762-767.

[5] J. Sun, J. Liu and L. Zhao, 2008. "Clustering Algorithms Research,". Journal of Software, vol. 19, no. 1, pp. 48–61.

[6] Bose, B.K., 1994. Expert system, fuzzy logic, and neural network applications in power electronics and motion control.

[7] Milovanovic B., Agatonovic M., Stankovic Z., Doncov N., 2012. Application of neural networks in spatial signal processing.

[8] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik, P. Arumugam, 2010. Financial Stock Market Forecast using Data Mining Techniques.

[9] S Abdulsalam Sulaiman Olaniyi, Adewole, Kayode S., Jimoh, R. G, 2010. Stock Trend Prediction Using Regression Analysis – A Data Mining Approach.