Financial Stock Market Forecast Using Data Mining in Palestine *

Dr. Eng. Yousef Abuzir **

Mr. AbdulRahman M.Baraka ***

Abstract:

This research involves building a model for forecasting stock price movements using data mining techniques with Artificial Intelligence classifier. Concerning the forecasting problem, we tackle the challenge of choosing Palestinian Stock Exchange (PSX) as an input data in the period between 2005-2017 and processing it accordingly to improve the predicting accuracy for the Palestinian stock market prices. We also address the issue of evaluating the prediction performance. Our results show that the best configuration for ANN is 6-5-1, which means six inputs, five neurons in hidden layer and one output. ANN classifier is feed-forward multi-layer perceptron neural network that is trained with back propagation algorithm. In addition, the results show we have a high degree of accuracy for prediction (0.010 +/- 0.000 for Root Mean Squared Error), this accuracy is higher than other related researches.

Keywords: ANN, Data Mining, Back Propagation Algorithm, Palestinian Stock Exchange.

التنبؤ بأسعار الأسهم في فلسطين

باستخدام تقنيات التنقيب في البيانات

ملخص:

في هذا البحث سوف نبني نموذجا للتنبؤ بحركات أسعار الأسهم باستخدام تقنيات التنقيب في البيانات مع مصنف الذكاء الاصطناعي. وفيما يتعلق بمشكلة التنبؤ، فإننا نتصدى للتحدي المتمثل في اختيار البورصة الفلسطينية كبيانات مدخلات في الفترة الزمنية (2005– 2017) ومعالجتها وفقا لذلك لتحسين دقة التنبؤ لأسعار سوق الأسهم الفلسطينية. كما نعالج مسألة تقييم أداء التنبؤ. وتبين نتائجنا أن أفضل تكوين للشبكة العصبية الذكية هو 6–5–1 مما يعني ستة مدخلات، وخمسة خلايا عصبية في الطبقة الخفية ومخرج واحد. وأن مصنف الشبكة العصبية الذكية هو عبارة عن شبكة عصبية متعددة الطبقات موجهة إلى الأمام يتم تدريبها عن طريق خوارزمية باك بروباغيشن. كذلك بينت النتائج أننا استطعنا الحصول على نسبة تنبؤ عالية حيث وصلت نسبة

الخطأ لمتوسط الجذر التربيعي إلى (+0.010/– 0.00) وهذه النسبة تفوق نسبة الدقة التي توصلت لها الأبحاث ذات العلاقة.

الكلمات المفتاحية: الشبكة العصبية الذكية، التنقيب في البيانات، خوارزمية باك بروباغيشن، بورصة فلسطين.

INTRODUCTION

With the emergence of information technology such as the World Wide Web and social networks, a large amount of data becomes available to all of us. Recently, data mining has received growing attention as a means to analyze and process a large amount of financial data (Berson et al, 2011) and (Leventhal, 2010). One of the challenging tasks for researchers, traders and investors is predicting stock price movements. Prediction of accurate prices of the stock presents a challenging task for all practitioners in stock market like traders and investors (Navale et al., 2016), (Huang et al., 2016). In general, the local and international stock market and company stock prices can be affected by many factors such as economic, political, war and civil unrest, natural disasters and terrorism, social and psychological. These factors shape, affect and interact with stock movement patterns.

In business and management, there is a need to understand the situations that occur in the stock markets domain to assist us in the analysis and investment decisions. In this study, we will discuss a research targeting stock markets domain to observe and record changes (events) when they happen, collect them and understand the meaning of each one of them. As well as, study most important factors and attributes that may affect stock prices and their integration in the classification.

We will build a model for forecasting stock price movements using data mining techniques with Artificial Intelligence classifier. Concerning the forecasting problem, we tackle the challenge of choosing Palestinian Stock Exchange (PEX) as an input data and process it accordingly to improve the predicting accuracy for the Palestinian stock market prices. We also address the issue of evaluating the prediction performance.

Prediction of stock market plays an important role in supporting financial decisions and stock business (Chen, 2009), (Dase & Pawar, 2010) and (Vaisla & Bhatt, 2010). The main purpose of this research is to investigate if and how data mining techniques can be used to predict the future trends of stock prices by analyzing financial data that is posted on the financial web pages.

There are many articles published on computational stock market prediction that uses numbers like stock price, volumes, company income, company cost, and so on, instead of textual information from news articles (Schumaker & Chen, 2010), (Schumaker & Zhang, 2012), (Pan & Poteshman, 2006) and (Robertson et al., 2007). It is important to investigate how stock markets react to breaking news because if we know this we can create fast-computerized systems that automatically analyze new news articles before the market has had time to adjust itself to the new information. Doing this opens up the possibility for making much more profit on stock trades and exchanges.

The results of this research will help managers of financial portfolios to understand the underlying factors behind the forecasting accuracy of stocks in the Palestinian Stock Exchange and other stock exchanges. This will further boost the confidence of stakeholders in the financial industry to do more business with less risk. It will provide companies and users with the ability to predict Palestinian stock market prices.

In this research, we aim to identify, investigate, analyze, evaluate and apply data mining techniques in stock movement in the Palestinian stock market (PEX, 2016) in order to predict expected stock prices.

The objectives of this research are as follows:

♦ Study existing Palestinian Exchange data on the official web site (PEX, 2016) for automatically analyzing financial data with the main focus on systems that use data mining methods for their prediction of future price trends.

♦ Study other external factors that may affect Palestinian market in particular.

♦ Investigate data mining methods that might

be used in an attempt to create an improved system.

♦ Design, implement and evaluate the proposed model that uses data mining techniques to generate automatically trading signals.

♦ Build a prediction classifier by using Artificial Intelligence technique (Fuzzy logic or Artificial Neural Network).

Marketing practitioners need forecasts to determine which new products or services to introduce or not; which markets to enter or exit; and which products to promote, sales people on the other hand, use forecasts to make sales plans that are generally based on estimates of future sales.

Hence having a good knowledge about share price movement in the future serves the interest of financial professionals and investors. This knowledge about the future boosts their confidence by way of consulting and investing. It goes without saying that forecasting methods which will predict the future movement of share prices with the least error margin will be of much interest to financial professional and investors.

There are many forecasting methods in projecting price movement of stocks. In this study, data mining methods will be employed in forecasting the price movement of stocks.

The main aim of this research is to build a model that forecasts stock price movements using data mining techniques. A case study and experiments will be presented and discussed based on applying data mining for the analysis of Palestinian Stock Exchange web. The performance of the experiments will be evaluated based on different data mining methods.

The rest of the paper is outlined as follows; a brief description about related researches. Section 3 presents the methodology used for experimenting the Back propagation (BP) algorithm. Then we analyzed and discussed our approach experiment setup, results and discussions. The final section represents the conclusion.

Literature Survey

The empirical study uses the variables of

technical analysis of stock market indicators for predicting stock market prices. A proposed model is tested and evaluated using PEX (PEX, 2016) data and the produced results showed a high level of accuracy in prediction.

Assaleh, K., El-Baz, H. and Al-Salkhadi, S. (Assaleh et al., 2011) present two prediction models for predicting stock prices. The first model was developed using back propagation feed forward neural networks. The second model was developed using polynomial classifiers (PC), as a first time application for PC to be used in stock prices prediction. The inputs to both models were identical, and both models were trained and tested on the same data. The study was conducted on Dubai Financial Market as an emerging market and applied on two of the market's leading stocks. In general, both models achieved very good results in terms of mean absolute error percentage. Both models showed an average error around 1.5% predicting the next day price, an average error of 2.5% when predicting second day price, and an average error of 4% when predicted the third day price.

Kunwar Singh Vaisla and Dr. Ashutosh Kumar Bhatt (Vaisla & Bhatt, 2010) proved that neural network (NN) outperform statistical technique in forecasting stock market prices. They have showed it through a method to forecast the daily stock price using neural network and then compared the result of the neural network forecast with the Statistical forecasting result. They have proved that neural network, when trained with sufficient data, proper inputs and with proper architecture, can predict the stock market prices very well. On the other hand, statistical technique's forecasting ability, though well built, is reduced as the series become complex. Therefore, NN can be used as a better alternative technique for forecasting the daily stock market prices.

Dase R.K. and Pawar D.D. (Dase & Pawar, 2010) tried to sum up the application of Artificial Neural Network for predicting stock market. They have also included that predicting stock index with traditional time series analysis. It has proven to be difficult. Artificial Neural network may be more suitable for the task. A neural network (NN) has the ability to extract useful information from large set of data. They have presented a review of literature about application of artificial neural network (ANN) for stock market predictions and from this literature; they have found that ANN is very useful for predicting world stock markets.

Pratyoosh Rai and Kajal Rai (Rai & Rai, 2016) have found from the comparison that the problem of stock index prediction is one of the most popular targets for various prediction methods in the area of finance and economics. In their article, the researchers have described the comparison of different neural network types for stock prediction. The prediction was carried out by modular neural network, ARIMA-based neural network, Genetic algorithm, Amnestic neural network, Multi-Branch neural network, etc. The authors have also performed comparative analysis of all these types of neural network (NN).

Few researches focused on Palestinian financial stocks market (Assaleh et al., 2011), (Al-Radaideh et al., 2013) and (Awad & Kmail, 2016), and these researches have many comments: first, the number of records in dataset that was used in there model is not described. Second, the period time of the selected data is not specified, especially that the Palestinian arena has a critical political periods that greatly have impact on the Palestinian stock exchange.

MATERIALS AND METHODS

The study population will be the investors, financial analyzers and researchers. We will use real data collected from PEX web site to build our dataset that shall be used in training and evaluating our prediction model. Most of the researches on Palestine financial markets depend on this data in their models and classifiers (Awad, 2016).

In general, other beneficiaries of the research are investors, shareholders, directors, regulators and other financial institutions as well as researchers in the academia, such as The Palestine Economic Policy Research Institute (MAS).

There are many different techniques used to build and evaluate stock markets analysis and prediction. In this study, we will use:

♦ Data Mining Methods to build a forecasting system for stocked market. Rapidminer tool

will be used with the different data mining methods to accomplish this research.

♦ Artificial Classifier for prediction, such as artificial neural network.

Also in this research, the exploratory research and systematic literature review will be used to analyze and apply the results of other researches, as follows:

♦ Collect and build dataset, that are utilized for training and predicting.

♦ Apply the various methods of data mining techniques to study and compare the performance of stock market movement from web related to Stock market to predict and assist in analysis and investment decisions.

♦ Feature Selection— In order to estimate the possible influence of each of the above attributes on the prediction.

♦ Classification— a process of finding a set of models that describe and distinguish data classes. This is done to achieve the goal of being able to use the model to predict the class which label is unknown. Many Classifiers can be used in this stage, such as the Artificial Neural Network or the Fuzzy Classifier.

♦ Researchers will use experimental research to evaluate the proposed methods and compare it with alternative approaches.

To achieve that, the research will consist of five stages: The first stage is to prepare the data and make it appropriate for the proposed model. Then, we will Analyze the data in graph to understand some data behavior. The third stage is to select attributes and study which attributes play a main role in prediction and the affected range. The next stage is to select the period time for dataset; determining which period time is suitable for prediction. The final stage is to build a proposed model with Back propagation Neural Network classifier (Dweib M. and Abuzir Y. 2017) by using Rapid Miner Studio. Many groups of experiments are applied on the proposed model to study the degree of impact for some attributes and parameters on prediction accuracy.

Preparing data

One year ago, for phase one, we collected

data from PEX for 2016-2017. In this phase, we suggest adding a new attribute called Month which may play a main role for enhancing the accuracy of the prediction. Therefore, we worked on collecting more data from previous years from PEX website (PEX, 2017), and we compiled data from the year 2005– until 2017. Bank of Palestine (BOP) presented us with 2868 records for seven attributes: Date, High price, Low price, Close price, Trade volume, Trades count and Trade value.

A previous Close price was added as an attribute to the dataset because it is an important attribute for prediction (Awad & Kmail, 2016).

All data types of attributes are integers or real except for the Date data type. We have encountered a problem with Date data type, because some records were provided as string types while others were produced as Date type. However, we extracted the month data type from the Date and added a new attribute to the dataset named Month and removed the Date attribute from the dataset.

Finally, we came up with a dataset that consists of 2868 records with eight attributes: Month, High price, Low price, Close price, Trade volume, Trades count, Trade value and Pre-close.

Data Analysis

The Palestinian region has experienced many political conditions, such as conflicts, wars, siege, etc., which affected the economic conditions of the country in general and the institutions in particular. The following chart (figure 1) illustrates the low value of the Close price of the envelopes in 2005-2017. We note that, in 2005, the value of Close price was approximately (5-6). However, in 2006 the stock price decreased to (2.6 - 3.1) because the region witnessed internal conflicts.

Although, trade counts increased in 2006 compared to 2005 despite the political situation as shown in figure 2.

Otherwise, stock price and trade count were stable, in average, from 2013 to 2017 as shown in figure 1 and figure 2. So, this period from 2013 may be appropriate for the current stock situation.

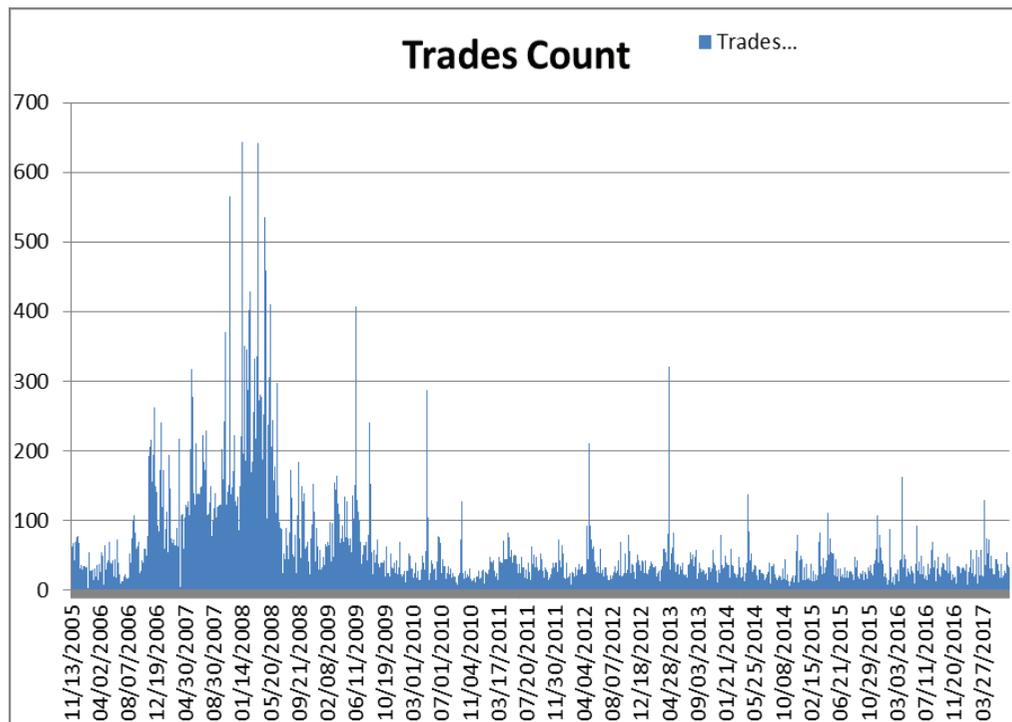**Figure 1: Close price value for period from 2005 to 2017.**

Figure 2: Trades Count value for period from 2005 to 2017.

Selecting attributes:

Commonly, related research use some related attributes. Awad and Kmail (2016) used High price, Low price, Close price and Trades count for their prediction model. Other research used attributes of High price, Low price, Close price, Previous Close and Open price (Al-Radaideh & Alnagi, 2013)

While Trade volume and Trade value depend on the value of trade count; we removed Trade value and Trade volume attributes from our dataset and only used Trade count.

Therefore, we came up with six attributes in dataset: Month, High price, Low price, Close price, Trades count and Previous close.

In the next stages, we will test the effect of each attribute on the accuracy of the prediction and then make the decision on whether to keep it or remove it.

Selecting the period time for dataset:

In the previous step, we showed that the stock market attributes started to stabilize from end of year of 2012. So, the period time after the year 2013 must be selected to deem its appropriateness to the current stock market.

Anyway, after building the model – in the next stage – we shall apply some experiments to test if old data enhances prediction accuracy or not.

Built Model:

Rapid Miner Studio version 7.6 was used to build our prediction model. Rapid Miner Studio is a powerful visual workflow designer for rapidly building predictive analytic workflows. This all-in-one tool features hundreds of data preparation and machine learning algorithms to support all your data science projects (Rapid Miner, 2017).

Our model has six steps as shown in figure 3. The first step consists of importing data from an excel sheet; then, selecting attributes that are used in the prediction. The third step is to replace missing value to a zero value, and finally split the dataset into two groups; the first group (75% of data) is used for learning stage and the second group (25% of data) is used for prediction stage. After that, we apply Artificial Neural Network (ANN) prediction model. While the final step is the validation, which tests the performance of classifier.
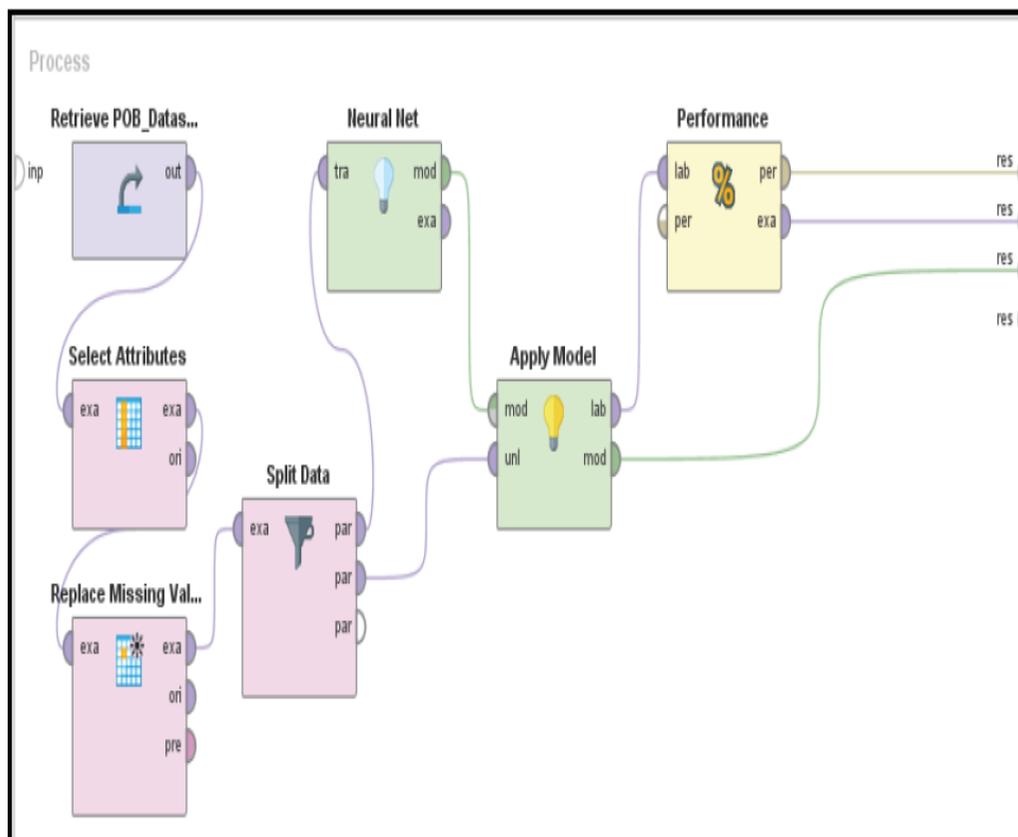

Figure 3: Prediction Model using Rapid Miner Studio.

ANN classifier is used in our model, the architecture of an ANN defines how its several neurons are arranged, or placed, in relation to each other. In general, an artificial neural network can be divided into three parts, named layers, which are known as:

(A) Input layer:

This layer is responsible for receiving information (data), signals, features or measurements from the external environment. These inputs (samples or patterns) are usually normalized within the limit values produced by activation functions. This normalization results in better numerical precision for the mathematical operations performed by the network.

(b) Hidden, intermediate or invisible layer:

These layers are composed of neurons which are responsible for extracting patterns associated with the process or system being analyzed. These layers perform most of the internal processing from a network.

(c) Output layer:

This layer is also composed of neurons, and thus is responsible for producing and presenting the final network outputs, which result from the processing performed by the neurons in the previous layers (Silva et al., 2017).

In addition, there are some parameters for Neural Network, which are:

- Training Cycles:

This parameter specifies the number of training cycles used for the neural network training. In Back propagation the output values are compared with the correct answer to compute the value of some predefined error-function. The error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. This process is repeated n number of times; where n can be specified using this parameter.

- Learning Rate:

This parameter determines how much we change the weights at each step. It should not be 0.

- Momentum:

The momentum simply adds a fraction of the previous weight update to the current one. This prevents local maxima and provides a smoothest optimization directions.

Figure 4 shown the architecture of Neural Network classifier that is used in our model and the parameters that are used, as follow:

♦ Input Layer: (5-6) attributes and one threshold.

♦ Hidden Layer.

♦ Output Layer.

♦ Training Cycles: 500.
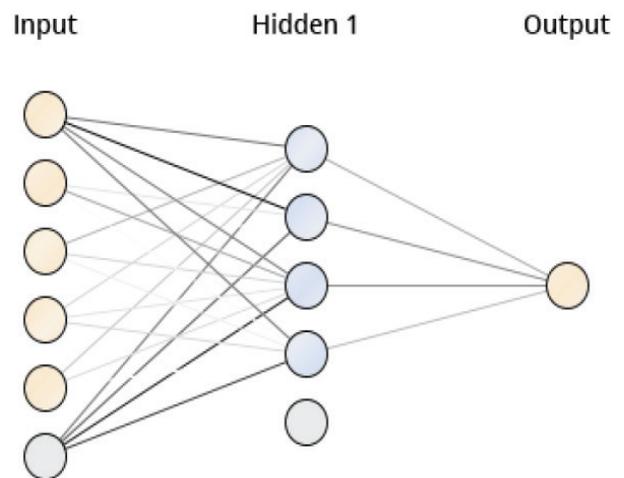
♦ Learning Rate: 0.3.

♦ Momentum: 0.2.

Figure 4: Architecture of Neural Network

Results and Discussion

We applied some groups of experiments to measure the effect of some attributes on prediction accuracy:

I. Group I:

This group measures the number of neurons in hidden layer of prediction accuracy. We use dataset with 634 records for this group.

Table 1:

Results of group I experiment.

Experiment No	Hidden Layer	Root mean squared error	Squared error
1	2	0.013 +/- 0.000	0.000 +/- 0.001
2	3	0.014 +/- 0.000	0.000 +/- 0.001
3	4	0.013 +/- 0.000	0.000 +/- 0.001
4	5	0.012 +/- 0.000	0.000 +/- 0.001
5	6	0.016 +/- 0.000	0.000 +/- 0.001

From the above results, we found that, the

best accuracy of the prediction with five neurons is in hidden layers.

II. Group II:

This group measures the time period effect on prediction accuracy. We use five neurons in hidden layers.

Table 2:

Results of group II experiment.

Exp. No	Period Time	Number Of Records	Root mean squared error	Squared error
1	2005-2017	2868	0.029 +/- 0.000	0.001 +/- 0.003
2	2013-2017	1152	0.012 +/- 0.000	0.000 +/- 0.000
3	2015-2017	634	0.012 +/- 0.000	0.000 +/- 0.001
4	2016-2017	261	0.010 +/- 0.000	0.000 +/- 0.000

From the above results we found that, the accuracy of the prediction increases as the period of time approaches the current period.

III. Group III:

This group measures the Month attribute effect on prediction accuracy.

Experiment 1: With Month attribute.

Table 3:

Results of group III-1 experiment.

Experiment No	Period Time	Root mean squared error	Squared error
1	2005-2017	0.029 +/- 0.000	0.001 +/- 0.003
2	2013-2017	0.012 +/- 0.000	0.000 +/- 0.000
3	2015-2017	0.012 +/- 0.000	0.000 +/- 0.001
4	2016-2017	0.010 +/- 0.000	0.000 +/- 0.000

Experiment 2: Without Month attributes.

Table 4:

Results of group III-2 experiment.

Experiment No	Period Time	Root mean squared error	Squared error
1	2005-2017	0.035 +/- 0.000	0.001 +/- 0.004
2	2013-2017	0.012 +/- 0.000	0.000 +/- 0.000

Experiment No	Period Time	Root mean squared error	Squared error
3	2015-2017	0.018 +/- 0.000	0.000 +/- 0.001
4	2016-2017	0.011 +/- 0.000	0.000 +/- 0.000

From the above results we found that, the prediction with the Month attribute is more accurate than the one without the Month attribute.

The previous experiment shows that the best accuracy we obtained was when we used six attributes: Month, High price, Low price, Close price, Trades count and the previous Price close, with hidden layer which consisted of five neurons. So, the best configuration for ANN is 6-5-1 which means six inputs, five neurons in hidden layer and one output. ANN classifier is feed-forward multi-layer perceptron neural network that is trained with Back propagation algorithm.

In addition, the accuracy ratio increases as the time period becomes closer to the current time period.

Figure 5 depicts the correlation level of accuracy by comparing the actual stock prices with the predicted values.
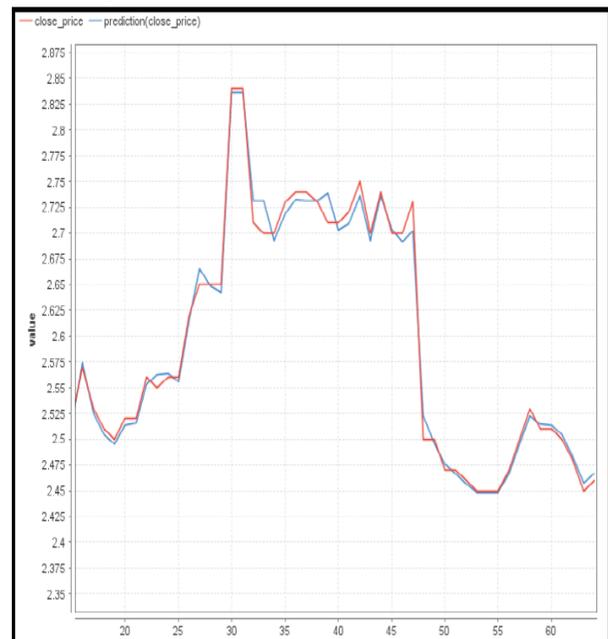

Figure 5: Comparing the actual stock prices with the predicted values.

Moreover, we got 0.010 +/- 0.000 of Root Mean Squared Error which means, the accuracy of our model is better than the accuracy of other related research (Awad and Kmail, 2016).

CONCLUSION

In this paper, we built a model for forecasting stock price movements in Palestine using data mining techniques. The main purpose of this research is to investigate how data mining techniques can be used to predict the future trends of stock prices by analyzing financial data that are posted on the Palestinian Stock Exchange web pages. This research showed that data mining techniques based on ANN can be used to improve the stock exchange prediction performance and it is possible to forecast stock price movements using data mining techniques on the Palestinian Stock Exchange web pages.

Our results show that the best configuration for ANN is 6-5-1 which means six inputs (Month, High-price, Low-price, Close-price, Trades-count and Previous-close), five neurons in hidden layer and one output. ANN classifier is feed-forward multi-layer perceptron neural network that is trained with Back propagation algorithm using Rapid Miner Tool. Our results showed a high level of accuracy in prediction. We obtained 0.010 +/- 0.000 of Root Mean Squared Error which means, the accuracy of our model is better than the accuracy of other related research.

Acknowledgement

References

1. Acciani, C., Fucilli, V., and Sardaro, R. (2011). Data Mining in Real Estate Appraisal: A Model Tree and Multivariate Adaptive Regression Spline Approach. Aestimum, 2011, Vol. 58, pp.27-45

2. Al-Radaideh Qasem A., Abu-Assaf Adel And Alnagi Eman. (2013). "Predicting Stock Prices Using Data Mining Techniques", The International Arab Conference on Information Technology (ACIT'2013).

3. Assaleh K., El-Baz H, and Al-Salkhadi (2011). Predicting Stock Prices Using Polynomial Classifiers: The Case of Dubai Financial Market. Journal of Intelligent Learning Systems and Applications. Vol 3, pp 82-89.

4. Awad M. and Kmail A. (2016). Towards Enhancing Stock Market Watching Based On Neural Network Predictions. International Journal of Information Technology and Business Management (JITBM). Vol.45 No.1.

5. Berson, A., Smith, S., and Thearling, K. (2011). An Overview of Data Mining Techniques [online]. [Accessed 28 November 2011]. Available at: http://www.thearling.com/text/dmtechniques/dmtechniques.htm

6. Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. Journal of Banking & Finance, 33(2):211–223.

7. Dase R.K. & Pawar D.D.(2010), "Application of Artificial Neural Network for stock market predictions: A review of literature", International Journal of Machine Intelligence, ISSN: 0975–2927, Volume 2, Issue 2, pp. 14-17.

8. Dweib M., Abuzir Y. (2017), Optimization of the Neural Networks Parameters, Palestinian Journal of Technology and Applied Sciences, pp 34 -47, No 1.

9. Huang,W., Nakamori, Y., andWang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32(10):2513–2522.

10. Leventhal, B. (2010). An introduction to data mining and other techniques for advanced analytics. Journal of Direct, Data and Digital Marketing Practice, pp. 137–153

11. Navale G. S., Dudhwala N. and Jadhav K. (2016). Prediction of Stock Market using Data Mining and Artificial Intelligence. International Journal of Computer

Applications, Volume 134 – No.12.

12. Palestine Exchange(PEX) Web site: http://www.p-s-e.com/psewebsite/english/Default.aspx

13. Pan, J. and Poteshman, A. M. (2006). The information in option volume for future stock prices. Review of Financial Studies, 19(3):871–908.

14. Qasem A. Al-Radaideh, Adel Abu-Assaf And Eman Alnagi. (2013). "Predicting Stock Prices Using Data Mining Techniques", The nternational Arab Conference on Information Technology (ACIT'2013).

15. Rai P. & Rai Kl. (2011), "Comparison of Stock Prediction Using Different Neural Network Types", International Journal of Advanced Engineering & Application.a

16. RapidMiner Studio. https://rapidminer.com/products/studio/.(10/8/2017).

17. RapidMiner Studio. https://rapidminer.com/products/studio/.(10/8/2017)

18. Robertson, C., Geva, S., & Wolff, R. C. (2007a). Can the content of public news be used to forecast abnormal stock market behaviour? In Proceedings of the Seventh IEEE International Conference on Data Mining (pp. 637–642). Washington, D.C.: IEEE Computer Society.

19. Schumaker, R. & Chen, H. (2010). A discrete stock price prediction engine based on financial news. Computer, 43(1), 51–56.

20. Schumaker, R., Zhang, Y., & Huang, C. (2012). Evaluating sentiment in financial news articles. Communications of International Information Management Association, 53, 458–464.

21. Silva N. Ivan, Spatti H.Danilo and others." Artificial Neural Network Architectures and Training Processes", Springer International Publishing Switzerland (2017).

22. Vaisla K. S. &. Bhatt A. K. (2010), "An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2104-2109.