Forecasting the Egyptian index movement of The Social Responsibility

Dr.Mamdouh Mowafy- Mona Eltaher

Statistics, Department, Faculty of Commerce, Ain Shams University, Egypt
Statistics, Department, Faculty of Commerce, Ain Shams University, Egypt

ABSTRACT

This paper illustrates the problem of predicting movement of The Companies' Social Responsibility index (S&P EGX) using historical data for 10 years in the form of daily data, applying on Artificial Neural Network (ANN) and Random Forest by using ten technical indicators as inputs to these models. This study divides S&P Index into segments by converting inputs from continuous to separate data, so separate form indicating the movement of the direction up or down based on their inherent properties. It focusses also on comparing the performance of these models in predicting when inputs are represented in real value form and specify direction of data. Where the study for both models, but Neural approved the efficiency of the classification network Model more accurate than Random forest Model.

Keywords:
Neural Network, Random Forest, Evaluating forecasts, Stock market, social responsibility.

Introduction

Private sector companies no longer depend on their profits or their financial positions. Modern concepts have emerged to help create a working environment capable of dealing with rapid economic, technological and managerial developments around the world.

The most prominent of these concepts was the concept of social responsibility of companies. The role of private sector institutions has become central to the development process, as demonstrated by the successes of the developed economies. Private sector institutions have realized that they are not isolated from society. They have become aware of the need to expand their activities to include more productive activities such as the concerns of society and the environment, the need to take in consideration the three pillars identified by the World Business Council for Sustainable Development: economic growth, social progress and environmental

protection, since it is difficult to predict the movement and value of stock market indices due to uncertainties. There are two types of analysis that the investor uses to make a purchase decision. The first type is the basic analysis, in which the investor looks to know the real value of stocks, industrial and economic conditions, political climate and others to determine he should invest or not. On the other hand, there is technical analysis, which evaluates stocks by studying the statistics resulting from stock market activity, such as stock prices and value. Technical analysis does not seek to measure the real value of stocks, but rather uses stock movement trends to determine the direction and behavior of stocks in the future.
There are many different approaches to predict index movement as in the Istanbul Stock Exchange **(Yakup, Melek and Ömer,2011)** aims to develop the efficiency of two models and compared their performances in predicting the direction of movement in the daily Istanbul Stock Exchange (ISE) National 100 Index. Models are based on classification techniques, artificial neural networks (ANN) and support vector machines (SVM). Ten technical indicators have been selected as inputs to the proposed models. Experiments were then conducted to detriment the parameter for both models to improve their prediction performance. The results showed that the average performance of the ANN model (75.74%) While **(Erkam, Gulgun and Tugrul, 2011)** was much better than the SVM model (71.52%). Based on neural network models that are effective in stock market projections. Models used are multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks using GARCH. All models were compared based on the Mean Square Error (MSE) and Mean Absolute Deviate (MAD). This was done using the daily average values of the Nasdaq Stock Exchange. **(Nair et al., 2011)** used the artificial neural network system to predict the value of the next day of the stock market index. The system is training on changing market movement with the help of a genetic algorithm that modifies the value of neural network parameters at the end of each trading session so that the best possible accuracy is obtained. The effectiveness of the system has been proven by five global stock indices using ten different performance metrics. The study by **(Hassan, Nath, and Kirley, 2007)** proposed and implemented a merger model between Hidden Markov Model (HMM), Artificial Neural Networks (ANN) and Genetic Algorithms (GA) to predict the behavior of the financial market. Using ANN Daily stock prices are converted into independent sets of values that become a contribution to HMM and depends on GA to improve the initial parameters of HMM. **(Chen, Leung, and Daouk, 2003)** forecast trend of Taiwan stock market index, one of the fastest growing stock exchanges in developing Asian countries. The study used the probabilistic neural network (PNN) to predict the direction of the index return after training by historical data, And measure the statistical performance of the PNN and compare it with Generalized methods of moments (GMM) Model. The study by **(Abraham, Nath and Mahanti, 2001)** used Hybridized soft computing techniques for stock market forecasting and trend analysis by utilizing the neural network for one day before stock forecasting and used the Neuro-Fuzzy system to analyze the expected stock price trend. Applied on the NASDAQ-100 index of Nasdaq Stock market, And there are two papers by **(Patel, J., Shah, S., Thakkar, P., & Kotecha, K. 2015)** addressed fusion of machine learning techniques to predict Stock market index, and applied on two indices from Indian markets. The first used hybrid models support vector regression- Artificial Neural Networks (SVR-ANN), SVR- (Random Forest)RF and SVR-SVR to compare with the single stage scenarios where ANN, RF and SVR are used single-handedly, ten technical indicators used as inputs to

each of the prediction models,. The second paper compare four prediction models, ANN. SVR, RF and naïve-Bayes with two approaches for input to these models.
This study focuses on comparing prediction performance of ANN and random forest by accuracy and f-measure to evaluate, using ten technical parameters as the inputs to these models to predict with S&P EGX, reflect the role of The Companies Social Responsibility in Egypt.

Research data

Ten years of data of S&P index value from 28 June 2007 to 29 March 2018. The data were obtained from <http://www.egyptse.com> website in the form of daily data and the calculation of the indicators of technical analysis based on previous study (Kara et al. 2011), which are ten technical indicators used as inputs for statistical models as table1 shows:

Table 1: Formulas for calculated indicators

Name of Indicators	Formulas
Simple Moving Average (n=10)	$\frac{C_t + C_{t-1} + \cdots + C_{t-9}}{n}$
Weighted day Moving Average(n=10)	$\frac{(10)C_t + (9)C_{t-1} + \cdots + C_{t-9}}{n + (n-1) + \cdots + 1}$
Momentum	$C_t - C_{t-9}$
Stochastic K%	$\frac{C_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} \times 100$
Stochastic D%	$\frac{\sum_{i=0}^{n=1} K_{t-i}}{10}\%$
Relative Strength Index (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} UP_{t-i/n})/(\sum_{i=0}^{n-1} DW_{t-i/n})}$
Moving Average Convergence Divergence (MACD)	$MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1})$
%Larry William's R	$\frac{H_n - C_t}{H_n - L_n} \times 100$
Exponential Moving Average	$EMA(k)_t = EMA(k)_{t-1} + \alpha \times (C_t - EMA(k)_{t-1})$
CCI (Commodity Channel Index)	$\frac{M_t - SM_t}{0.015D_t}$

Where:

C_t: Closing price

at the specific period L_t: lowest price

H_t: Highest price at the specific period

$DIFF_t$=EMA (12 periods)$_t$-EMA (26 periods)$_t$

Smoothing factor $\alpha = \frac{2}{K+1}$

K: period Time for Exponential Arithmetic average

/ HH_t: Lowest low and highest high in the last t days LL_t

$$M_t = \frac{H_t + L_t + C_t}{3}, SM_t = \frac{(\sum_{i=1}^{n} M_{t-i+1})}{n},$$

$$D_t = \frac{(\sum_{i=1}^{n} |M_{t-i+1} - SM_t|)}{n}$$

DW_t: Downward price change UP_t: Upward price change

Table (2) shows the highest and lowest values, mean and the standard deviation of the variables used. S&P represents the value of the S&P index while the S&P Dummy variable indicates the change in the index if a positive change takes value 1 and if a negative change takes the value 0, to use the statistical methods in the classification.

Table 2: Descriptive Statistics for variables

Variable	Valid N	Mean	Minimum	Maximum	Std.Dev.
S&P index	2574	1183.959	548	2806.00	439.974
S&P Dummy variable	2574	0.526	0	1.00	0.499
ıple Moving Average (SM	2574	1181.008	557	2725.00	434.232
‹ponential Moving Averag (EMA)	2574	1181.021	558	2726.38	433.740
›ving Average Convergen‹ Divergence (MACD)	2574	1171.620	8	2623.21	420.968
Weighted day Moving Average (WMA)	2574	1180.035	554	2714.09	432.817
Momentum (MOM)	2574	5.864	-337	361.00	69.090
Commodity Channel Index	2574	-21.874	-255667	48000.00	5216.299

(CCI)					
Stochastic K% (STCK)	2574	56.684	0	100.00	38.479
Stochastic D% (STCD)	2574	56.648	1	100.00	28.192
Relative Strength Index (RSI)	2574	47.480	-78699	22221.21	1983.278
(WR)%Larry William's R	2574	43.316	0	100.00	38.479

Prediction Models

Artificial Neural Networks (ANN) model

Inspired by the performance of biological neural networks, artificial neural networks are a dense network of interlocking neurons that are activated based on inputs. In this study, a neural network consisting of three layers will be used where nutrition is forward. The network inputs are ten technical indicators represented by 10 neurons in the input layer, while the output layer consists of one neuron with the log sigmoid function.
Since these results are related values between 0 and 1, the mean 0.5 is used to determine the outcome of predicting the movement of the indicator up or down. If the resulting value is greater than or equal to 0.5, the prediction is that the movement of the indicator is higher take value 1. Both neurons in the hidden layer contain the log sigmoid conversion function.
The output of the network is compared to the real output to calculate the error in the prediction, followed that return the error to the networks to modified the weights of the connection the training the data again and repeated the previous steps several times until the network output is obtained very close to the real outputs This reduces the error, while using sample test to avoid overfitting.

Random forests

Learning tree to make decision is one of the most popular techniques for classification. Accuracy of classification is not comparable with other classification methods, and it is very effective. It represents the evolution of decision tree techniques. Random forests belong to the ensemble learning algorithms. The decision tree is used as a base leaner of ensemble.
The idea of ensemble learning is that one classification is not enough to determine a class of testing data. The reason is, based on the sample data, the classifier is unable to distinguish between the noise and the pattern of data, So it performs sampling with replacement, After its establishment when testing the data used, the decision which the majority of trees come with, it is considered the final product, this also avoids the problem of over-training.
The algorithm for the implementation of random forest is as follows:
Input: D group training, the number of trees in the ensemble K
Outputs: Composite Model

The number of trees in the ensemble trees is the parameters of the model and for its efficient identification, the range is 10-200 with an increase of 10 at a time during parameter preparation experiments.

Experimental results

Summery of Training Model

Artificial Neural Networks

The Active neural networks *Summary is explained by* Table3, type MLP with 10 inputs, 13 neurons in the hidden units, and 2 outputs (corresponding to the two categories of the target percentage of the sample size of the training to test is (7:3), with a train, variable *S&P index*), test and validation classification rate at 85.0166%, 83.4196% and 81.088%, respectively. the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is an iterative method for solving unconstrained nonlinear optimization problems was used to train the network and the best solution was found at training cycle number 199. The network has a Tanh activation for the hidden units and Softmax activation function for the outputs (and hence the use of the Entropy error function).

: Summary of ANN٣ Table

Training perf.	**85.01665**
Test perf.	83.41969
Validation perf.	81.08808
Training algorithm	BFGS 199
Error function	Entropy $[Hy'(y):=-\sum iy'i\log(yi)]$
Hidden activation	Tanh $[tanh(x)=2\cdot\sigma(2x)-1]$
Output activation	Softmax

Random Forest

Demonstrates the basic mechanism of how the *Random Forest* algorithm implemented can avoid overfitting illustrated in (Figure 1).

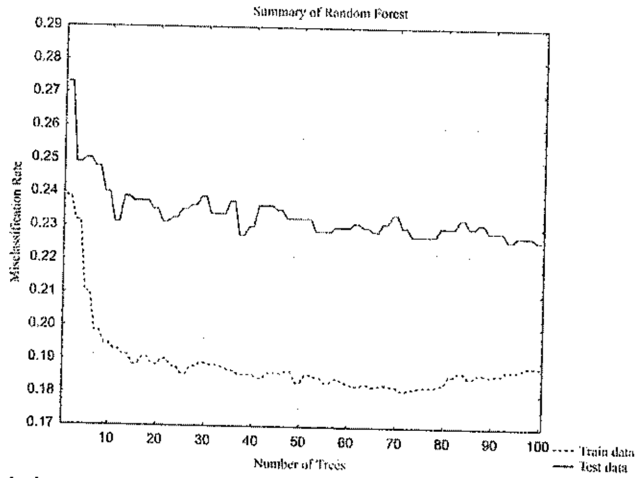

Fig. 1. Summary of training random forest.

In general, as more and more simple trees are added to the model, the misclassification rate for training data will generally decrease. The same trend should be observed for misclassification rates defined over the testing data. However, as more and more trees are added the misclassification rate for the testing, at one point the data will start to increase (while the misclassification rate for the training set keeps decreasing), clearly the marking point where evidence for overfitting is beginning to show, so the training stopped at 100 trees with train and test Risk Estimate rate at 0.188446 and 0.226296 respectively, as table 4 show:

Table 4: Risk estimates Response for random forest

	Risk Estimate	Standard error
Train	0.188446	0.009261
Test	0.226296	0.014878

Evaluation Chart

Artificial Neural Networks

Used ROC curve to evaluate the goodness of fit for S&P classifier as shown (Figure 2) It is a plot of the true positive rate against the false positive rate for the different possible cut points. A ROC curve demonstrates the trade-off between sensitivity and specificity. Also helps to assess model performance. The greater area under the curve was 0.937888, so it means good model performance

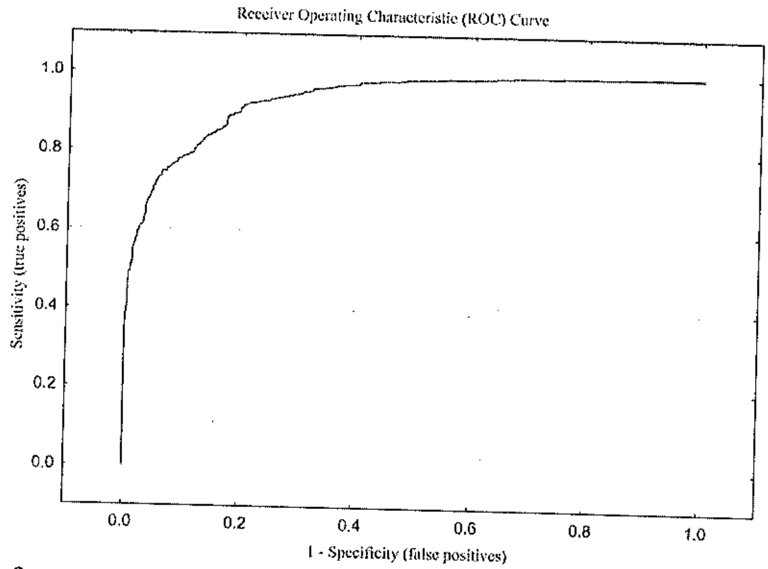

Fig. 2. Roc Curve for ANN.

Random Forest

The 3D histogram of the classification matrix between the Predicted vs. observed illustrate the efficiency of classification, the overall correct classification is 81.16%, (1/1) 82.02% and (0/0) 80.21% compared with incorrect classification as shown in (figure 3)

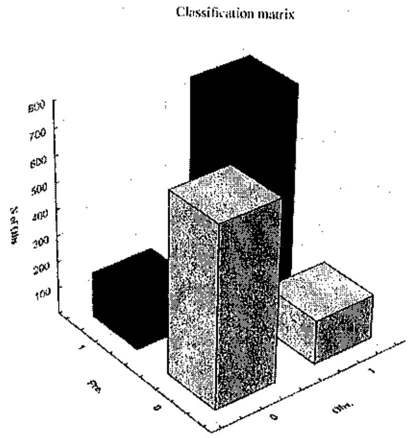

Fig. 3. Classification Matrix for random forest.

Predictor importance

This bar Chart (Figure4) of the predictor importance, to determine the variables, make the major contributions to the prediction of the dependent variable of interest display a bar chart that pictorially shows the importance ranking on a 0-1 scale for each predictor variable considered in the analysis. The most important indicators of technical analysis that affect The Egyptian index of The Companies Social Responsibility are Stochastic K%, and Larry William's R% with more than 90% then Momentum and Stochastic D% nearly 60%, after that Relative strength and commodity channel index with average between 40% to 50% while rest of variable less than 40%.

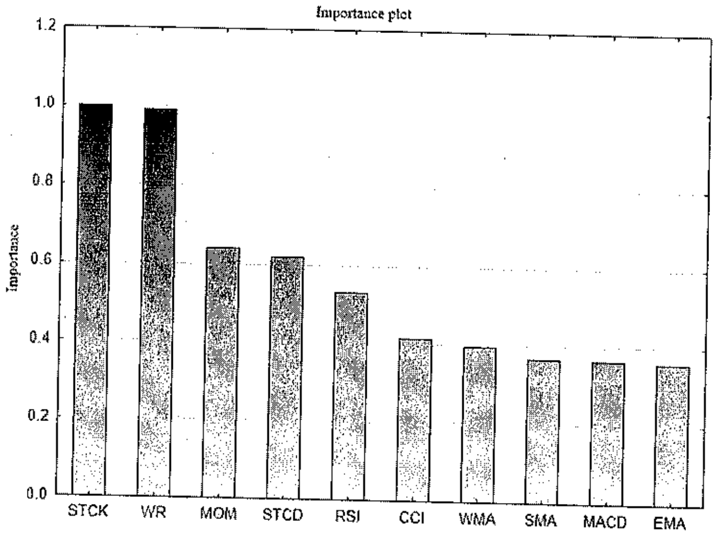

Fig. 4. Predictor importance for random forest.

Evaluation measures

Using accuracy and F-measure to evaluate the performance of models, by calculating Precision and Recall which are evaluated from True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) getting value using these eqs.

$$Precision_{Positive} = \frac{TP}{TP+FP} \quad (1)$$

$$Precision_{Negative} = \frac{TN}{TN+FN} \quad (2)$$

$$Recall_{Positive} = \frac{TP}{TP+FN} \quad (3)$$

$$Recall_{Negative} = \frac{TN}{TN+FP} \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

$$F-measure = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (6)$$

Precision calculated by the weighted average of precision positive and precision negative and the same for Recall is the weighted average of recall positive and recall negative.

Table 5: Models Performance

	Accuracy	Precision	Recall	F-measure
ANN	0.850166	0.850166	0.8501	0.850133106
Random Forest	0.811554	0.811554	0.811574	0.811563729

Table 5 reports accuracy and f-measure of each models, random forest model exhibits less performance with 81.11% accuracy. Neural network model is slightly high with 85.013% accuracy.

Conclusions

The task focused in this paper on predicting direction of movement for The Companies Social Responsibility index. Prediction performance of two models namely ANN and random forest is compared based on ten years (2007-2018) of historical data of S&P EGX index. Ten technical parameters used to learn each one of these models.
The above evidence shows that, Technical indicators have a high ability to predict direction of movement for The Companies Social Responsibility index, and the most important of these indicators are Stochastic K%, Larry William's R%, Momentum and Stochastic D%. for both models, but Neural network Experiment approved the efficiency of the classification Model more accurate with average 85% nearly, than Random forest Model which is nearly 81%.

The importance of predicting S&P EGX index is the importance of focusing on the role of companies towards social responsibility, ensuring to a certain extent the support of all members of the society for their goals and development mission, recognizing their existence and contributing to the success of their objectives. So prospective studies can be done on how to increase the role of The Companies Social Responsibility.
While previous studies have proven the efficiency of statistical methods based on machine learning compared to traditional statistical methods. This study compares two methods of machine learning to identify those with higher prediction performance in this field. In the future studies, it is possible to focus on the inclusion of other methods of machine learning and comparing them in this field and others.

References:

Abraham, A., Nath, B., & Mahanti, P. K. 2001. Hybrid Intelligent Systems for Stock Market Analysis. Computational Science - ICCS, Springer, pp. 337-345.

Byun, H., & Lee, S.-W. 2002. Applications of support vector machines for pattern recognition: A survey, in Pattern recognition with support vector machines. Ed: Springer, pp. 213-236.

Caruana, R., & Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. 23rd International Conference on Machine Learning.

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

Cortes, C., & Vapnik, V. 1995. Support-vector networks. Machine learning, vol. 20, pp. 273-297.

. Empirical analysis of model selection criteria for Garg, A., Sriram, S., & Tai, K. 2013 . IEEE conference.genetic programming in modeling of time series system

Gino J. Lim & Eva K. Lee, 2008. Optimization in Medicine and Biology. New York: Auerbach Publications, p. 592.

Guresen, E., Kayakutlu, G., Daim, T.-U. 2011. Using artificial neural network models in stock Pages 10389–10397.،market index prediction. Expert Systems with Applications, 38

Hassan, M. R., Nath, B., & Kirley, M. 2007. A fusion model of HMM, ANN and GA for stock market forecasting. Elsevier, Expert systems with Applications,33, pp. 171–180.

. Forecasting stock market movement direction Huang, W., Nakamori, Y., &Wang, S.-Y. 2005 with support vector machine. Computers & Operations Research, 32, pp. 2513–2522.

Karaa, Y., Acar Boyacioglub, M., & Baykanc, Ö. K. 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Elsevier, Expert Systems with Applications, 38, pp. 5311–5319.

Khemchandani, R., Chandra, S., et al., 2009. Knowledge based proximal support vector machines. European Journal of Operational Research, pp. 914–923.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. 2015. Predicting stock and stock price index movement using trend deterministic Data preparation and machine learning techniques. Expert Systems with Applications, 42, pp. 259–268.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. 2015. Predicting stock market using fusion of machine learning techniques. Elsevier, Expert Systems with Applications, 42, pp. 2162–2172.

Richard O. Duda, Peter E. Hart, and David G. Stork, 2000. Pattern Classification. Wiley editorial, 2nd Ed.

Predicting stock returns by classifier Tsai, C.-F., Lin, Y.-C., Yen, D.-C., & Chen, Y.-M. 2011. . Applied Soft Computing, Volume 11, Issue 2, pp.2452-2459. ensembles

Vapnik, V. N. 1999. An overview of statistical learning theory. IEEE Transactions on Neural Networks, pp. 988–999.

Xu, X., Zhou, C., & Wang, Z. 2009. Credit scoring algorithm based on link analysis ranking with support vector machine. Expert Systems with Applications, 36, pp. 2625–2632.

Software:

STATISTICA (data analysis software system), version 12. StatSoft, Inc. (2014). www.statsoft.com.

التنبؤ بحركة المؤشر المصري للمسئولية الاجتماعية

الملخص

يوضح هـذا البحث التنبؤ بحركة مؤشر المسؤولية الاجتماعية للشـركات (S&P EGX) باستخدام البيانات التاريخية لمدة ١٠ سنوات فـي شكـل بيانات يومية، والتطبيق علـى الشبكة العصبية الاصطناعية (ANN) والغابـات العشوائية Random Forest باستخدام عشرة مؤشرات فنية كمدخلات لهذه النماذج. وتقـوم هذه الدراسة بتقسيم البيانات الى قطاعات بتحويل المدخلات من بيانات مستمرة إلى منفصلة. حيث ان كل معلمات الإدخال تكون في شكل منفصل يشير الى حركة الاتجاه لأعلى أو لأسفل على أساس الخصائص الملازمة لها. وينصب التركيز أيضا على مقارنة أداء هذه النماذج في التنبؤ عندما يتم تمثيل المدخلات فـي شكل قيم حقيقية وتحديد اتجاه البيانات. واثبتت الدراسة كفاءة التصنيف لكلا النموذجين، ولكـن نموذج الشبكات العصبية أكثر دقة من نموذج الغابة العشوائية.

الكلمات الدالة :

الشبكات العصبية، الغابات العشوائية، كفاءة التنبؤ، سوق الأوراق المالية، المسؤولية الاجتماعية.