

Stock Price Prediction Using *K*-Nearest Neighbor (*k*NN) Algorithm

Khalid Alkhatib¹

Hassan Najadat²

Ismail Hmeidi³

Mohammed K. Ali Shatnawi⁴

Abstract

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.

Keywords: stock price prediction, listed companies, data mining, k-nearest neighbor, non-linear regression.

1. Introduction

Recent business research interests concentrated on areas of future predictions of stock prices movements which make it challenging and demanding. Researchers, business communities, and interested users who assume that future occurrence depends on present and past data, are keen to identify the stock price prediction of movements in stock markets (Kim, 2003). However, financial data is considered as complex data to forecast and or predict. Predicting market prices are seen as problematical, and as explained in the efficient market hypotheses (EMH) that was put forward by Fama (1990). The EMH is considered as bridging the gap between financial information and the financial market; it also affirms that the fluctuations in prices are only a result of newly available information; and that all available information reflected in market prices. The EMH assert that stocks are at all times in equilibrium and are difficult for inventors to speculate. Furthermore, it has been affirmed that stock prices do not pursue a random walk and stock prediction needs more evidence (Gallagher and Taylor, 2002; Walczack, 2001; Kavussanos and Dockery, 2001; Lakonishok et.al, 1994; O'Connor et. al., 1997; Lo and MacKinlay, 1997; Kirt and Malaikah, 1992; Lo and MacKinlay, 1988). Moreover, various studies were performed to determine stock price predictions (Subha and Nambi, 2012; Qian and Rasheed, 2007; Fama and French, 1992; Cochrane, 1988; Campbell, 1987; Chen, et al. 1986; Basu, 1977).

In addition to purchasing and selling stocks and shares in stock markets, each stock is not only characterized by its price, but also by other variables such as closing price which represents the most important variable for predicting next day price for a specific stock. There is a relationship and specific behavior exists between all variables that effect stock movements overtime. Different economic factors, such as political stability, and other unforeseeable circumstances are variables that have been considered for stock price predictions (Ou, P. and Wang, H., 2009; Fama and French, 1993; Cochrane, 1988; Campel, 1987; Chen. et.al.1986). Table 1 summarizes the main variables that affect stock movements used in this article.

^{1,3,4} Department of Computer Information Systems, Jordan University of Science and Technology, Irbid, Jordan

² Department of Computer Science, College of Computers and Information Technology, Taif University Taif, KSA, and currently on sabbatical leave from Jordan University of science and Technology, Jordan

Data mining technology is used in analyzing large volume of business and financial data, and it is applied in order to determine stock movements. Mining temporal stock markets is required to provide additional capabilities required in cases where the existing data and their interactions need to be observed through time dimension.

In stock predictions, a set of pure technical data, fundamental data, and derived data are used in prediction of future values of stocks. The pure technical data is based on previous stock data while the fundamental data represents the companies' activity and the situation of market. Combining data mining classification approaches in stock prediction yields a future value for each unknown entities of companies' stocks values based on historical data. This prediction uses various methods of classification approaches such as neural networks, regression, genetic algorithm, decision tree induction, and k-Nearest Neighbors (*k*NN). In classification approaches, a data set is divided into training data set and testing set. *k*NN uses similarity metrics to compare a given test entity with the training data set. Each data entity represents a record with *n* features. In order to predict a class label for unknown record, *k*NN selects *k* recodes of training data set that are closest to the unknown records.

The rest of the article will be structured as follows; section 2 will exemplify the review of the relevant literature, while section 3 describes the research methodology used and analysis. Section 4 shows the data description, results and analysis. The Non-linear regression results are included in section 5. Finally the conclusion is seen in section 6.

2. Literature Review

Financial services companies are developing their products to serve future prediction. There are a large amount of financial information sources in the world that can be valuable research areas, one of these areas is stock prediction and also called stock market mining. Stock prediction becomes increasingly important especially if number of rules could be created to help making better investment decisions in different stock markets.

The genetic algorithm had been adopted by Shin et al. (2005); the number of trading rules was generated for Korea Stock Price Index 200 (KOSPI 200), in Sweden Hellestrom and Homlstrom (1998) used a statistical analysis based on a modified *k*NN to determine where correlated areas fall in the input space to improve the performance of prediction for the period 1987-1996. Both models mentioned were provided in the Zimbabwe stock exchange to predict the stock prices which included Weightless Neural Network (WNN) model and single exponential smoothing (SES) model Mpofo (2004). Clustering stocks approach was provided by Gavrilov et al. (2004) to group 500 stocks from the Standard & Poor. The data represented a series of 252 numbers including the opening stock price. A fuzzy genetic algorithm was presented by Cao (1977) to discover pair relationship in stock data based on user preferences. The study developed potential guidelines to mine pairs of stocks, stock-trading rules, and markets; it also showed that such approach is useful for real trading. Moreover, other studies adopted *k*NN as prediction techniques such as (Subha et al., 2012; Liao et al. 2010; Tsai and Hsiao 2010; Qian and Rasheed, 2007)

3. Research Methodology And Analysis

The *k*NN algorithm method is used on the stock data. Also, mathematical calculations and visualization models are provided and discussed below.

3.1 k-Nearest Neighbor Classifier (*k*NN)

K-nearest neighbor technique is a machine learning algorithm that is considered as simple to implement (Aha et al. 1991). The stock prediction problem can be mapped into a similarity based classification. The historical stock data and the test data is mapped into a set of vectors. Each vector represents *N* dimension for each stock features. Then, a similarity metric such as Euclidean distance is computed to take a decision. In this section, a description of *k*NN is provided. *k*NN is considered a lazy learning that does not build a model or function previously, but yields the closest *k* records of the training data set that have the highest similarity to the test (i.e. query record). Then, a majority vote is performed among the selected *k* records to determine the class label and then assigned it to the query record.

The prediction of stock market closing price is computed using *k*NN as follows:

- a) Determine the number of nearest neighbors, *k*.
- b) Compute the distance between the training samples and the query record.

- c) Sort all training records according to the distance values.
- d) Use a majority vote for the class labels of k nearest neighbors, and assign it as a prediction value of the query record.

3.2 Mathematical Calculations and Visualizations Models

This represents an overview of equations that were applied in this article for predicting next day price. The calculations includes error estimation, total sum of squared error, average error, cumulative closing price when sorted using predicted values, k-values and training Root Mean Square (RMS) errors.

- a) Root Mean Square Deviation (RMSD) is accuracy metric that computes the differences between the estimated values, Y, and the actual values, X. The total of RMSD is aggregated into a single value measure. $RMSD = \sqrt{\sum (Y-X)^2}$.
- b) Explained Sum of Squares (ESS) is computed as follows:

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 .$$

Where y_i is the predicted variable, and y is the actual value.

- c) Average Estimated Error (AEE)

AEE is the total sum of RMS errors for all variables in stock records divided by the total number of the records.

$$AEE = \frac{\sum_{i=1}^n RMS_i}{n}$$

3.3 Visualization Graph

To evaluate the performance of kNN learning model, lift graph is applied and drawn for different companies' stock values. The lift chart symbolizes the enhancement that a data mining model offers when distinguished against a random estimation, and the change is expressed in terms of lift score. Through contrasting the lift scores for a variety of parts of the data set and for different models, it can then be decided which model is supreme and which percentage of the cases within the data set would gain from employing the predictions model. Furthermore, using the lift chart assist in distinguishing how accurate predictions are for various models with identical predictable characteristic. The lift graph also shows the ratio between the results obtained using the predictive model or not. The other graph applied is the plot curves to show the relation between the actual and predicted stock price.

4. Data Description, Results, And Analysis

In this article, data from the Jordanian stock exchange was analyzed and a brief data analysis is presented to provide the reader with the fundamental concepts of data attributes. Also, the obtained results of prediction of the Jordanian stock exchange are provided.

4.1 Data Description

The sample data was extracted from the Jordanian stock exchange. The study sample included stock data of five randomly selected companies listed on the Jordanian stock exchange as a sample training dataset from the period June 4, 2009 to December 24, 2009 as shown in table 2. Each of these companies has approximately 200 records with three attributes including closing price, low price, and high price as shown in table 3. A brief data analysis is presented with the fundamental concepts of data attributes. The attributes for each company are included in the data analysis. Closing price is the main factor that affects the prediction process for a specific stock based on kNN algorithm. The kNN algorithm is applied on a 1000 records to estimate predicted values for each stock.

4.2 Analysis And Results

The results of the predicted stock price for each individual company used in the sample with graphs for the actual and predicted prices are presented.

The results as seen in tables 4.1 to 4.5 and in figures 1 to 11 are those after applying kNN algorithm for each company's closing prices with the residual values which indicates how far away is the predicted values from the actual values; the negative residual value indicates that the predicted value is larger than the actual one.

Section 4.3 summarizes the five companies' prediction performance. Tables 4.1 to 4.5 respectively represent the results after applying k NN algorithm of the Arab international for education and investment (AIEI), Jordan steel company (JOST), Arab financial investment (AFIN), Irbid district electricity (IREL), and the Arab potash company (APOT).

As depicted in the figures (1-10) below, the line chart of the actual and predicted values for the companies in the sample and after adopting the k NN prediction model, the results show that the predictive value and the actual value were moving in similar manner as seen in figure 2,4,6,8, and 10. Moreover, the lift chart also applied to evaluate the performance of k NN learning model used and proved that the model used is performing well; this can also be seen in figures 1,3,5,7, and 9 representing the lift charts (company's dataset).

4.3 Prediction Performance Evaluations

Table 6 represents a summary of the total squared errors, RMS errors and the average errors for the five companies. The residuals offer the differences between the predicted values and actual the values in the sample data. The table also shows that the values of errors are very small which indicate that the actual value and predicted value are close. This yields a high accuracy of using the k NN algorithm in predicting stock values.

5. Non-Linear Regression Results

Non-linear regression is a data analysis technique in which the observed data is incorporated into a model presented in a mathematical non-linear function combining the model parameters that relies on independent variables used. GraphPad Prism v5.02 software was used to apply centered second order polynomial (quadratic) non-linear regression which has the following formula:

$$\text{Price} = B_0 + B_1 (\text{day} - \text{mean}(\text{day})) + B_2 (\text{day} - \text{mean}(\text{day}))^2$$

Where:

B_0 , B_1 and B_2 : Constants.

Day: Actual day in which we will predict the price.

Price: Predicted price depending on the day.

Figure 11 provides a graphical representation of non-linear regression. Also, the figure shows a computed regression equation for each company. These companies are AIEI, JOST, APOT, AFIN, and IREL. Based on these equations, investors can compute the future values of stock prices for each company.

6. Conclusion

In this paper, a prediction process for five listed companies on the Jordanian Stock Market was carried out, and is considered to be the first of its type implemented in Jordan as a case study using real data and market circumstances. Consequently, a robust model was constructed for the purpose set out. The data was extracted from five major listed companies on the Jordanian stock exchange, the sample data was used to be our training data set (about 200 records for each company) upon the criteria previously mentioned to apply our model. We adopted an efficient prediction algorithm tool of k NN with $k=5$ to perform such tests on the training data sets we had. According to the results, k NN algorithm was stable and robust with small error ratio, so the results were rational and reasonable. In addition, depending on the actual stock prices data; the prediction results were close to actual prices. Having such rational results for predictions in specific, and for using data mining techniques in real life; this presents a good indication that the use of data mining techniques could help decision makers at various levels when using k NN for data analysis. So, we consider that employing this prediction model, k NN is real and viable for stock predictions. Nevertheless, the implementation and the full make use of information systems technology and due to the lack of knowledge of financial econometrics in Jordan, there is still long way to utilize advanced predicting models to help the financial markets and brokerage houses and to move forward and be part of the developed international financial markets. Furthermore, this may weakens the attractiveness of investments in the Jordanian market which eventually weakens the market return. The study also shows that contemporary data mining techniques offer the world of finance useful stock market movements' prediction analysis.

References

- Aha, D., Kibler, D.W., Albert, M.K. (1991). Instance-based learning algorithms. *Mach Learn*, 6, 37–66
- Alexander, S.S. (1961). Price movements in speculative markets: Trends or random walks. *Ind Manage Rev*, 7–26
- Breiman, L. (1996). Bagging predictors. *Mach Learn* 24(2), 123–140
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A. (1984). *Classification and regression trees*. Chapman & Hall (Wadsworth, Inc.). New York
- Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D* 110, 43–50
- Cootner, P.H. (1964). *The random character of stock market prices*. MIT Press, MA
- Corazza, M., Malliaris, A.G. (2002). Multi-fractality in foreign currency markets. *Multinat Fin J* 6(2), 65–98
- Dietterich, T.G. (1997). Machine-learning research: Four current directions. *AI Magazine* 18(4), 97–136
- Dietterich, T.G. (2000). *Ensemble methods in machine learning*. First International Workshop on Multiple Classifier Systems. New York
- Fama, E.F. (1965). The behaviour of stock market prices. *J. Bus* 38, 34–105
- Fama, E.F. (1991). Efficient capital markets. II *J. Fin.* 46(5), 1575–1617
- Fama, E.F., Fisher, L., Jensen, M., Roll, R. (1969). The adjustment of stock price to new information. *Int. Eco Rev* 10(1), 1–21
- Frank, R.J., Davey, N., Hunt, S.P. (2000). Input window size and neural network predictors. *IEEE-INNS-ENNS Int Joint Conf Neural Netw (IJCNN'00)*. 2, 2237–2242
- Gallagher, L., Taylor, M. (2002). Permanent and temporary components of stock prices: Evidence from assessing macroeconomic stocks. *Southern Eco J* 69, 245–262
- Gately, E. (1996). *Neural networks for financial forecasting*. Wiley, New York
- Grech, D., Mazur, Z. (2004). Can one make any crash prediction in finance using the local Hurst exponent idea? *Physica A. Statistical Mech Appl.* 336, 133–145
- Hagan, M.T., Demuth, H.B., Beale, M.H. (1996). *Neural network design*. PWS Publishing, Boston, MA
- Hagan, M.T., Menhaj, M. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Netw.* 5(6), 989–993
- Hansen, L., Salamon, P. (1990). Neural network ensembles. *IEEE Trans Patt Analy Mach Intell.* 12, 993–1001
- Hellstrom, T., Holmstrom, K. (1998). Predicting the stock market. Technical report series IMA-TOM-1997-07. Center of Mathematical Modeling, Malardalen University
- Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feed forward networks are universal approximators. *Neural Net* 2(5), 259–366
- Hsieh, D.A. (1991). Chaos and nonlinear dynamics: application to financial markets. *J Fin* 46, 1839–1877
- Hurst, H.E. (1951). Long-term storage capacity of reservoirs: an experimental study. *Trans Amer Soc Civil Engi* 116, 770–799
- Jensen, M.C. (1978). Some anomalous evidence regarding market efficiency. *J Fin Eco* 6, 95–102
- Kavussanos, M.G., Dockery, E. (2001). A multivariate test for stock market efficiency: The case of ASE Applied Financial Economics. 11(5), 573–579(7)
- Kirt, C.B., Malaikah, S.J. (1992). Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia. *J Bank & Fin* 16 (1), 197–210
- Lo, A.W., MacKinlay, A.C. (1997). Stock market prices do not follow random walks. *Market Efficiency: Stock Market Behavior in Theory and Practice* 1, 363–389
- Mandelbrot, B. (1982). *The fractal geometry of nature*. W.H. Freeman. New York
- Mandelbrot, B.B., Ness, J.V. (1968). Fractional brownian motions, fractional noises and applications. *SIAM Rev* 10, 422–437
- May, C.T. (1999). *Nonlinear pricing: theory & applications*. Wiley, New York
- Peters, E.E. (1991). *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*. Wiley, New York
- Peters, E.E. (1994). *Fractal market analysis: applying chaos theory to investment and economics*. Wiley, New York
- Qian, B., Rasheed, K. (2004). Hurst exponent and financial market predictability. *Proceedings of the 2nd IASTED international conference on financial engineering and applications*. Cambridge, MA, USA, 203–209
- Soofi, A.S., Cao, L. (2002). *Modelling and forecasting financial data: techniques of nonlinear dynamics*. Kluwer Academic Publishers: Norwell, Massachusetts
- Walczak, S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. *J Manag Infor Syst* 17,4, 203–222

Table 1: Stock market variables that affect investor decisions in buy/sell transactions

Variable	Description
Price	Current price
Opening Price	Opening price for a specific trading day
Closing price	Closing price for a specific trading day
High	Highest price in a specific day
Low	Lowest price in a specific day

Table 2: Companies listed on the Jordanian stock market used in the sample

Company Symbol	Description	Number of records
AIEI	The Arab International For Education & Investment.	200
JOST	Jordan Steel company	200
AFIN	Arab Financial Investment	181
IREL	Irbid District Electricity	189
APOT	The Arab Potash	200

Table 3: Variables used

Variable Name	Description
Closing price	Current price for a stock
Low price	Lowest price in a specific day for a stock
High price	Highest price in a specific day for a stock

Table 4.1: The results after applying *k*NN algorithm for the (AIEI)

Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual
2.85	2.9	0.05	2.624	2.63	0.006	2.505	2.55	0.045
2.825	2.8	-0.025	2.64	2.64	0	2.505	2.5	-0.005
2.825	2.85	0.025	2.69	2.69	0	2.505	2.5	-0.005
2.9	2.9	0	2.703333	2.68	-0.023333	2.505	2.5	-0.005
2.9	2.9	0	2.568	2.62	0.052	2.505	2.5	-0.005
2.993333	3.03	0.036667	2.505	2.5	-0.005	2.505	2.5	-0.005
3	3	0	2.443333	2.4	-0.043333	2.505	2.5	-0.005
2.91	2.91	0	2.32	2.32	0	2.576364	2.55	-0.026364
2.703333	2.78	0.076667	2.42	2.42	0	2.505	2.5	-0.005
2.703333	2.65	-0.053333	2.45	2.45	0	2.576364	2.55	-0.026364
2.576364	2.67	0.093636	2.568	2.54	-0.028	2.576364	2.55	-0.026364
2.6	2.6	0	2.443333	2.48	0.036667	2.593333	2.59	-0.003333
2.505	2.5	-0.005	2.45	2.45	0	2.576364	2.55	-0.026364
2.555	2.55	-0.005	2.443333	2.45	0.006667	2.505	2.5	-0.005
2.555	2.6	0.045	2.505	2.5	-0.005	2.576364	2.56	-0.016364
2.576364	2.59	0.013636	2.568	2.54	-0.028	2.505	2.55	0.045
2.61	2.61	0	2.568	2.54	-0.028	2.576364	2.55	-0.026364
2.624	2.6	-0.024	2.51	2.51	0	2.56	2.56	0
2.61	2.61	0	2.42	2.42	0	2.505	2.5	-0.005

Note: AIEI predicted closing prices after applying *k*NN algorithm. 200 records from the period Jan 4, 2009 to Dec 24, 2009 are selected as the training dataset and only 57 records are shown in the table. Total squared RMS error is 0.263, RMS error is 0.0378 and the average error is -5.434E-09

Table 4.2: The results after applying *k*NN algorithm for Jordan steel company (JOST)

Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual
2.47	2.47	0	2.706667	2.73	0.023333	2.64	2.68	0.04
2.473333	2.47	-0.003333	2.69	2.69	0	2.59	2.59	0
2.47	2.47	0	2.703333	2.73	0.026667	2.672	2.65	-0.022
2.5325	2.52	-0.0125	2.672	2.68	0.008	2.703333	2.72	0.016667
2.615	2.58	-0.035	2.703333	2.66	-0.043333	2.755	2.7	-0.055
2.64	2.61	-0.03	2.69	2.69	0	2.82	2.82	0
2.64	2.61	-0.03	2.72	2.76	0.04	2.82	2.83	0.01
2.64	2.65	0.01	2.672	2.65	-0.022	2.962	2.9	-0.062
2.63	2.62	-0.01	2.63	2.64	0.01	2.9275	2.9	-0.0275
2.63	2.64	0.01	2.706667	2.69	-0.016667	2.923333	2.92	-0.003333
2.63	2.62	-0.01	2.79	2.79	0	2.83	2.83	0
2.645	2.65	0.005	2.672	2.71	0.038	2.8725	2.85	-0.0225
2.615	2.65	0.035	2.516667	2.59	0.073333	2.95	2.94	-0.01
2.64	2.63	-0.01	2.473333	2.47	-0.003333	2.9275	2.95	0.0225
2.64	2.7	0.06	2.516667	2.47	-0.046667	2.896667	2.9	0.003333
2.64	2.6	-0.04	2.516667	2.49	-0.026667	2.8725	2.93	0.0575
2.675	2.67	-0.005	2.5325	2.51	-0.0225	2.923333	2.88	-0.043333
2.645	2.6	-0.045	2.5325	2.51	-0.0225	2.98	2.95	-0.03

Note: 200 records from the period Jan 4, 2009 to Dec 24, 2009 are chosen as the training dataset and only 57 records are shown in the table. Total squared RMS error is 0.263, RMS error is 0.0378 and the average error is -5.4347E-09

Table 4.3: The results after applying *k*NN algorithm for Arab financial investment (AFIN)

Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual
3.396667	3.39	-0.006667	3.31	3.25	-0.06	3.36	3.4	0.04
3.338	3.39	0.052	3.31	3.25	-0.06	3.384286	3.39	0.005714
3.431667	3.4	-0.031667	3.24	3.24	0	3.36	3.32	-0.04
3.396667	3.4	0.003333	3.23	3.23	0	3.37	3.43	0.06
3.431667	3.4	-0.031667	3.28	3.28	0	3.386667	3.44	0.053333
3.413333	3.4	-0.013333	3.356667	3.36	0.003333	3.384286	3.33	-0.054286
3.431667	3.4	-0.031667	3.32	3.32	0	3.4	3.4	0
3.431667	3.4	-0.031667	3.245	3.25	0.005	3.413333	3.4	-0.013333
3.49	3.46	-0.03	3.33	3.27	-0.06	3.396667	3.4	0.003333
3.49	3.46	-0.03	3.338	3.3	-0.038	3.384286	3.4	0.015714
3.49	3.46	-0.03	3.295	3.31	0.015	3.384286	3.35	-0.034286
3.338	3.4	0.062	3.356667	3.31	-0.046667	3.413333	3.44	0.026667
3.24	3.24	0	3.338	3.3	-0.038	3.465	3.45	-0.015
3.245	3.24	-0.005	3.33	3.39	0.06	3.384286	3.4	0.015714
3.24	3.24	0	3.338	3.3	-0.038	3.44	3.44	0
3.31	3.3	-0.01	3.295	3.28	-0.015	3.465	3.48	0.015

Note: AFIN predicted closing prices after applying *k*NN algorithm. 181 records from the period Jan 4, 2009 to Dec 24, 2009 are chosen as the training dataset and only 54 records are shown in the table. Total squared RMS error is 0.263, RMS error is 0.036 and the average error is -1.005E-08

Table 4.4 The results after applying *k*NN algorithm for Irbid district electricity (IREL)

Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual
9.15	9.15	0	8.35	8.35	0	7.58	7.58	0
8.72	8.72	0	8.45	8.5	0.05	7.444	7.22	-0.224
8.803333	8.81	0.006667	8.1	8.1	0	7.574615	7.5	-0.074615
9.026667	9	-0.026667	8.17	8.17	0	7.444	7.78	0.336
9.28	9.28	0	8.6	8.6	0	7.41	7.41	0
9.2	9.2	0	8.45	8.4	-0.05	7.8	7.8	0
9.026667	9.08	0.053333	7.995	8.01	0.015	7.792	7.6	-0.192
8.803333	8.8	-0.003333	8.4	8.4	0	7.8	7.8	0
8.803333	8.8	-0.003333	8.29	8.29	0	7.792	7.6	-0.192
9.026667	9	-0.026667	8.095	7.9	-0.195	8	8	0
9.24	9.24	0	8.095	8.29	0.195	7.574615	7.74	0.165385
9.1	9.1	0	7.825	7.95	0.125	7.346667	7.38	0.033333
9.4	9.4	0	7.8	7.8	0	7.03	7.03	0
8.87	8.99	0.12	7.574615	7.6	0.025385	7.433333	7.4	-0.033333
8.87	8.75	-0.12	7.87	7.87	0	7.71	7.71	0
8.51	8.51	0	7.574615	7.5	-0.074615	7.574615	7.5	-0.074615
8.56	8.55	-0.01	7.574615	7.5	-0.074615	7.42	7.42	0
8.56	8.57	0.01	7.8	7.8	0	7.07	7.07	0

Note: AFIN predicted closing prices after applying *k*NN algorithm. 189 records from the period Jan 4, 2009 to Dec 24, 2009 are chosen as the training dataset and only 54 records are shown in the table. Total squared RMS error is 1.282, RMS error is 0.1046 and the average error is 4.2735E-08

Table 4.5: The results after applying *k*NN algorithm for The Arab potash (APOT)

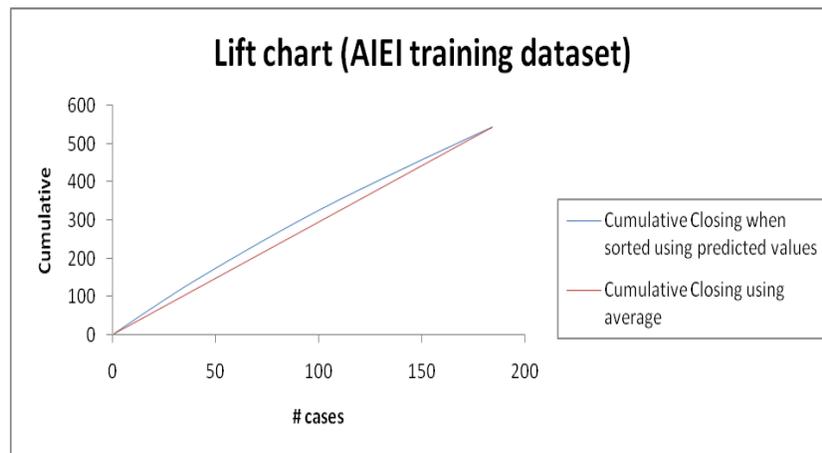
Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual	Predicted Value	Actual Value	Residual
33.145	34.18	1.035	32.76	32.76	0	29.97	29.97	0
33.82	33.51	-0.31	34.29	34.48	0.19	30.69	30.69	0
33.77	34	0.23	34.533333	34.05	-0.483333	30.34	30.34	0
34	34	0	34.01	34.01	0	30.36	30.36	0
34.49	34.49	0	33.82	33.95	0.13	31.945	31.9	-0.045
34.98	34.98	0	33.77	33.54	-0.23	32.25	31.5	-0.75
35.166667	34.87	-0.296667	34.833333	34.25	-0.583333	33.145	32.7	-0.445
34.75	34.75	0	34.4	34.4	0	33.65	33.3	-0.35
34.99	34.99	0	33.525	33.51	-0.015	32.85	32.85	0
34.6	34.6	0	33.525	33	-0.525	33.11	33.11	0
34.29	34.1	-0.19	33.145	33	-0.145	34.61	34.61	0
34.825	34.45	-0.375	33	33	0	34.84	34.84	0
34.39	34.39	0	33.12	33.12	0	33.525	33.6	0.075
34.29	34.29	0	31.6	31.6	0	32.9	32.9	0
34.533333	34.4	-0.133333	30.88	30.88	0	31.91	31.91	0
34.3	34.3	0	29.9	29.9	0	31.945	31.99	0.045
34.435	33.97	-0.465	30.4	30.4	0	32.9	32.9	0

Note: APOT predicted closing prices after applying *k*NN algorithm. 200 records from the period Jan 4, 2009 to Dec 24, 2009 are chosen as the training dataset and only 51 records are shown in the table. Total squared RMS error is 22.7453, RMS error is 0.337 and the average error is 2.49E-08

Table 6: Prediction performance evaluations for the whole sample companies (five companies) after applying *k*NN algorithm for *k*=5

<i>k</i> NN algorithm for <i>K</i> = 5			
Company	Total squared RMS error	RMS error	Average error
AIEI	0.263151	0.0378176	-5.43E-09
AFIN	0.2629177	0.0363482	-1.01E-08
APOT	22.74533	0.3372338	2.50E-08
IREL	1.2823397	0.1046908	4.27E-08
JOST	0.17963	0.0300444	1.508E-08

**Figure 1
Lift chart for AIEI training dataset**



Note: The area between the baseline and the curve is an indicator of the goodness of the model.

Figure 2 : Plot graph shows the relationship between AIEI’s predicted/actual closing price for 1 year period

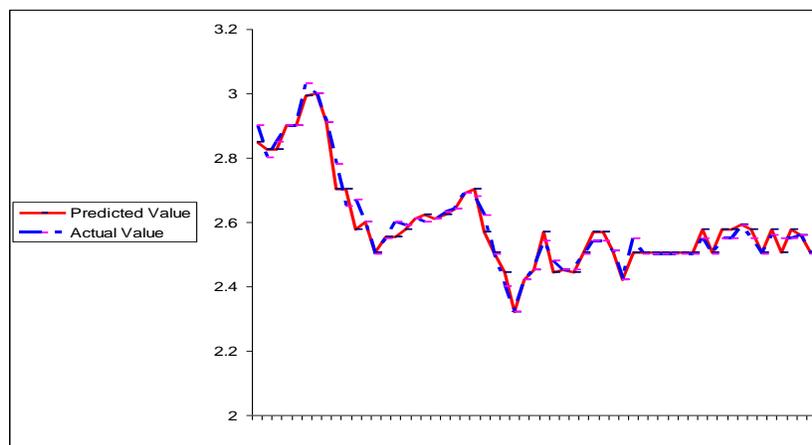
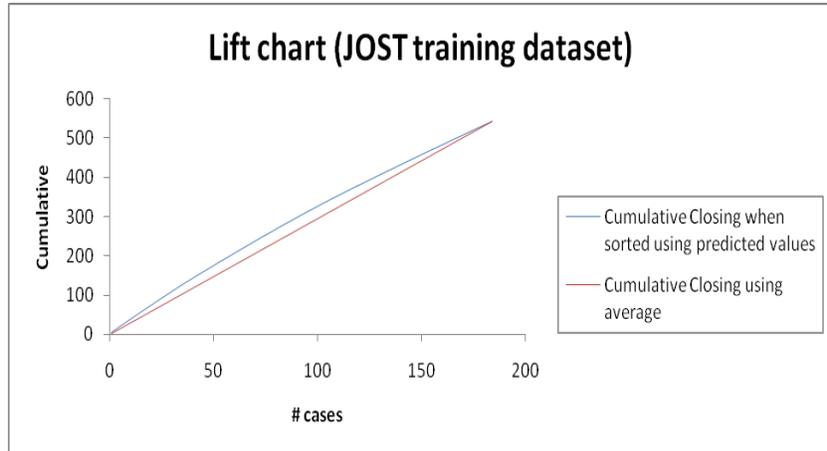


Figure 3: Lift chart for JOST training data set



Note: The area between the baseline and the curve is an indicator of the goodness of the model.

Figure 4: Plot graph shows the relationship between JOST's predicted/actual closing price for 1 year period

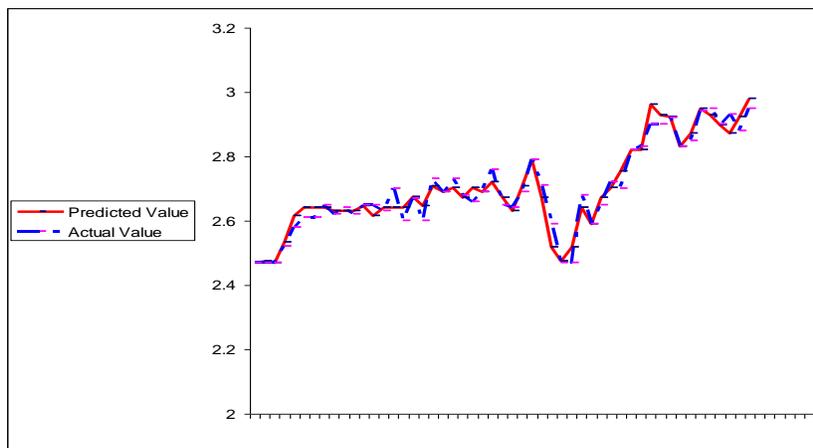
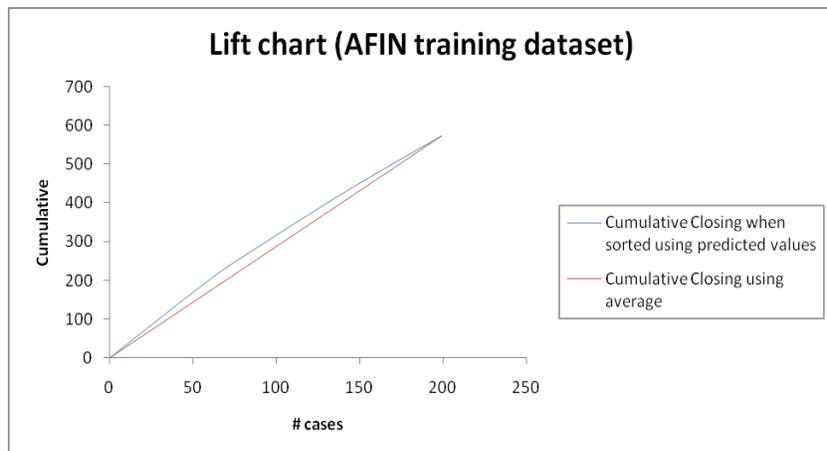


Figure 5: Lift chart for AFIN company training data set



Note: The area between the baseline and the curve is an indicator of the goodness of the model.

Figure 6: Plot graph shows the relationship between AFIN’s predicted/actual closing price for 1 year period

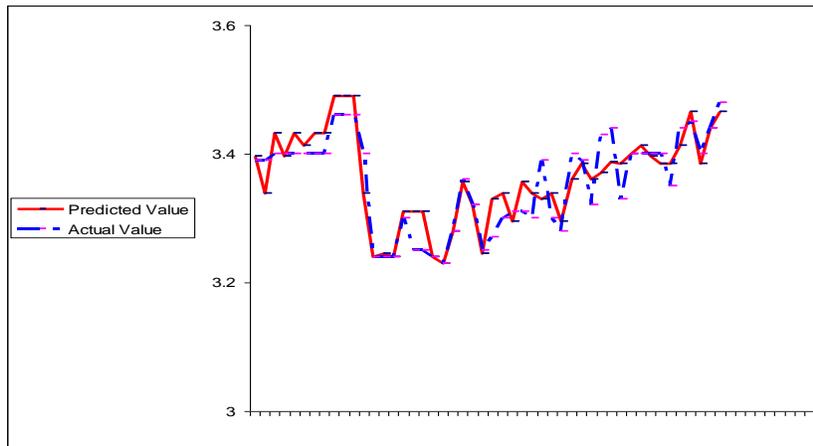
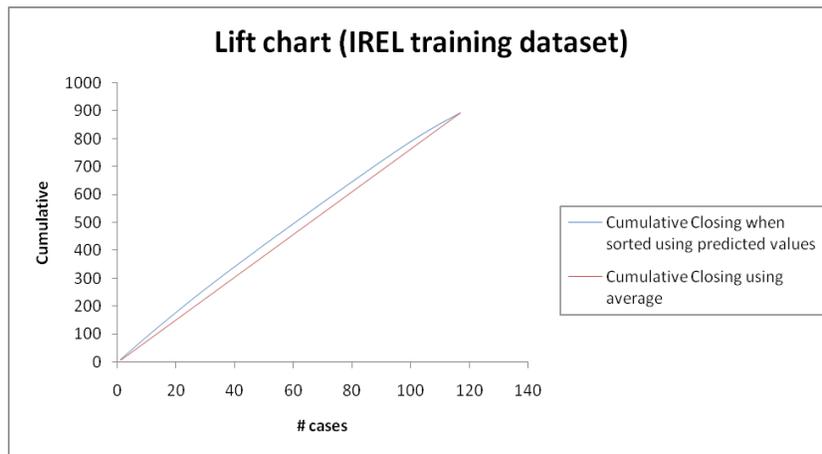


Figure 7: Lift chart for IREL company training data set



Note: The area between the baseline and the curve is an indicator of the goodness of the model.

Figure 8: Plot graph shows the relationship between IREL’s predicted/actual closing price for 1 year period

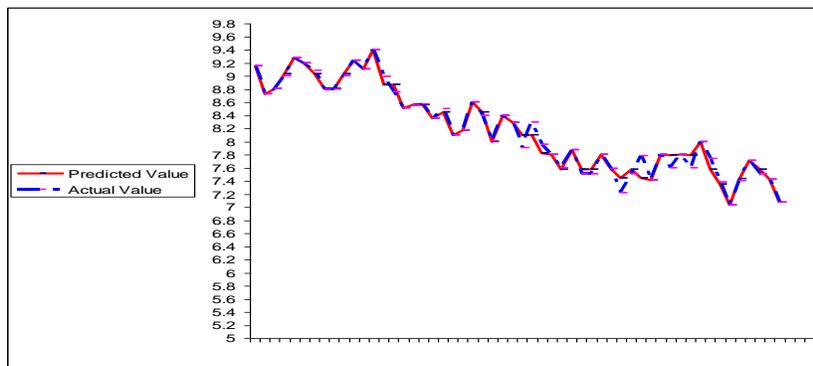
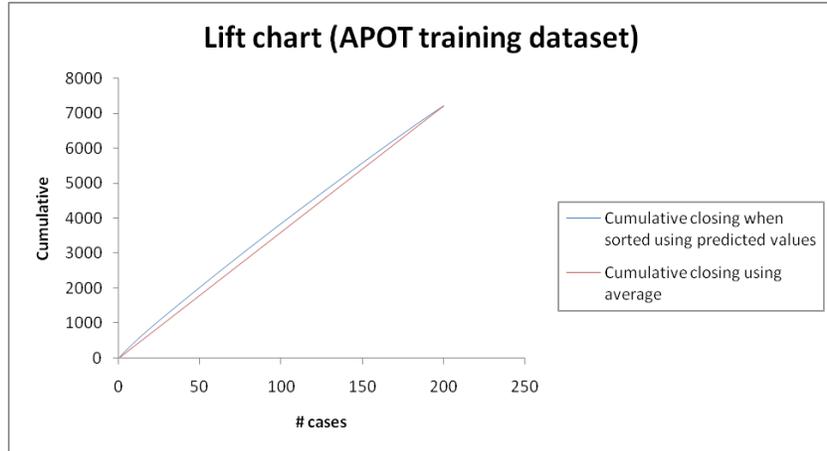


Figure 9 : Lift chart for APOT company training data set



Note: The area between the baseline and the curve is an indicator of the goodness of the model.

Figure 10: Plot graph shows the relationship between APOT's predicted/actual closing price for 1 year period.

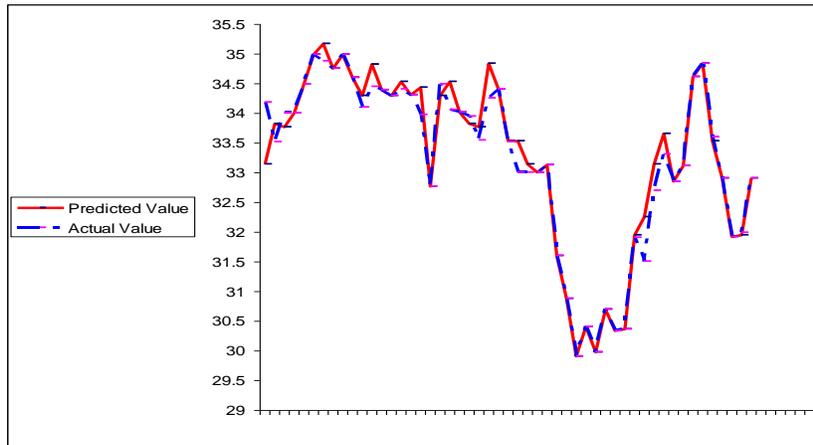


Figure 11: Regression distributions for the sample companies AIEI, JOST, APOT, AFIN, and IREL.

