Journal of Economics, Management and Trade

24(6): 1-13, 2019; Article no.JEMT.51284
ISSN: 2456-9216
(Past name: British Journal of Economics, Management & Trade, Past ISSN: 2278-098X)

Predicting Daily Returns of Global Stocks Indices: Neural Networks vs Support Vector Machines

Jasleen Kaur[1*] and Khushdeep Dharni[1]

[1]School of Business Studies, Punjab Agricultural University, Ludhiana, Punjab, India.

Authors' contributions

This work was carried out in collaboration between both the authors. Author JK carried out the study by collecting and analyzing the data. She also prepared the first draft of the manuscript. Author KD guided and managed the tasks of designing the study and statistical analysis. He contributed in the finalization of the manuscript. Authors have read and approved the final manuscript.

Article Information

DOI: 10.9734/JEMT/2019/v24i630179
Editor(s):
(1) Kamarulzaman Ab. Aziz, Director, Entrepreneur Development Centre, Multimedia University, Persiaran Multimedia, Malaysia.
Reviewers:
(1) Sachin Kamley, India.
(2) Der-Jang Chi, Chinese Culture University, Taiwan.
Complete Peer review History: http://www.sdiarticle3.com/review-history/51284

Original Research Article

Received 19 June 2019
Accepted 04 September 2019
Published 13 September 2019

ABSTRACT

Uniqueness in economies and stock markets has given rise to an interesting domain of exploring data mining techniques across global indices. Previously, very few studies have attempted to compare the performance of data mining techniques in diverse markets. The current study adds to the understanding regarding the variations in performance of data mining techniques across the global stock indices. We compared the performance of Neural Networks and Support Vector Machines using accuracy measures Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) across seven major stock markets. For prediction purpose, technical analysis has been employed on selected indicators based on daily values of indices spanning a period of 12 years. We created 196 data sets spanning different time periods for model building such as 1 year, 2 years, 3 years, 4 years, 6 years and 12 years for selected seven stock indices. Based on prediction models built using Neural Networks and Support Vector Machines, the findings of the study indicate there is a significant difference, both for MAE and RMSE, across the selected global indices. Also, Mean Absolute Error and Root Mean Square Error of models built using NN were greater than Mean Absolute Error and Root Mean Square Error of models built using SVM.

**Corresponding author: E-mail: jazz831@gmail.com;*

Keywords: Stock market forecasting; neural networks; support vector machines.

1. INTRODUCTION

Capital is an important means of production and stock market plays a crucial role in mobilizing capital for various business activities including food processing, textiles, fertilizers and pesticides etc. Stock market plays a significant role in the economic growth of the country to a great extent. Stock market is a place where public listed company's shares are traded. The variations of stock market depend on variations of numerous indicators representing the agriculture, industry and service sector. Therefore, stock market returns are affected by various factors in these sectors. Stock markets generates enormous amount of complex and non-linear data. One of the most challenging tasks in modern finance is to find an efficient way to analyze stock market data so as to provide investors useful information for investment decisions. The purpose of prediction is to reduce uncertainty associated with investment decision making. There are multifarious methods available to deal with such an enormous amount of data. But, due to inherent limitations of traditional forecasting techniques in building a model to predict the future values accurately, data mining techniques took prominent place in the domain of stock market prediction. The major drawbacks to traditional methods are: incorrect number of variables, incorrect forecasting model and incorrect values of coefficients of these parameters. These issues can be solved using data mining techniques. In data mining, model is built iteratively till the extraction of unknown patterns and relationships in the data which are almost inconceivable by human imagination.

A large body of research on application of data mining to stock market has been produced. Data mining techniques can effectively deal with the nonlinearity of the stock market and allows a search for valuable information, in large volumes of data [1]. For making profitable trades, investors are highly interested in forecasting the future trend of stock market indices and stock prices. Further, uniqueness in economies and stock markets has given rise to an interesting domain of exploring data mining techniques across global indices. The current study adds to the understanding regarding the variations in performance of data mining techniques across the global stock indices.

Present paper aims to compare the performance of data mining techniques across global stock markets by using the popular techniques such as Neural Networks and Support Vector Machines. Comparisons are also drawn between NN and SVM in terms of MAE and RMSE of predicted values of daily returns.

1.1 Background of the Study

Data mining has established itself as a theoretically sound alternative to traditional statistical models in stock market study. Data mining technique is a science and technology of exploring data in order to discover previously unknown patterns and is a part of the overall process of Knowledge Discovery in Databases (KDD). Data mining is a powerful tool for information extraction from large volumes of data [2]. These techniques have become an increasingly important research area [3,4,5].

Applications of data mining techniques encompasses wide variety of domains including credit card use [6], customer relationship management [7], bankruptcy prediction [8,9], bacteriology for bacterial identification [10], MIG welding process [11], detecting blog spam [12], fault diagnosis and condition monitoring [13,14], software fault prediction [15], machining parameter optimization [16], demand forecasting [17], emotional speech analysis [18] and software engineering [19]. Data mining techniques have been used in a wide range of stock market prediction applications. This range includes stock price forecasting, stock index forecasting and forecasting stock prices with the help of external factors [20,21,22,23,24].

The preference of Neural networks is quite evidence in the literature due to its the accuracy in terms of direction of prediction [25,26,27,28, 29,30]. On the other hand, SVM has also been preferred by many researchers [31,32,33,34].

Testing the performance NN and SVM across global stock indices is a relatively newer research domain [5]. Outcome of the study will shed the light on utility of NN and SVM for predictive modeling in stock indices across the globe.

2. METHODS

The present manuscript explores the performance of NN and SVM across global stock indices. Selection of stock indices, data collection and procedures adopted for carrying out the study are given in this section.

2.1 Sampling and Data Collection

On the basis of Morgan Stanley Capital International (MSCI) market classification, we selected seven countries across the globe for the study [35] which includes three developed markets (United States, United Kingdom and Japan) and four emerging markets (China, Brazil, India and South Africa). Further, we selected indices from largest stock exchanges of these countries on the basis of turnover of financial derivative segment. Description of the selected indices is provided in Table 1.

Closing, Opening, High and Low values of selected stock indices were recorded for the period of twelve years starting from 1st April, 2005 to 31th March, 2017.

2.2 Data Transformation

The existence of continuous, noisy and complex data may pose a challenging task to extract information from the raw data [36]. Therefore, data is transformed to improve the predictive power of the techniques [32,37,38]. We used 12 technical indicators [32,38] for predicting direction of stock indices. These indicators include Stochastic %K, Stochastic %D, Stochastic Slow %D, Momentum, rate of change (ROC), Larry Willaim's %R (LW %R), A/D Oscillator (Accumulation/Distribution), Disparity 5-days, Disparity 10-days, OSCP(Price Oscillator), CCI (Commodity Channel Index) and RSI (Relative Strength Index). These indicators are elaborated in Table 2.

We created data sets spanning different time periods for model building such as 1 year, 2 years, 3 years, 4 years, 6 years and 12 years. Therefore, we made 28 data sets for each index. The data sets were created for all seven indices. In all 196 data sets have been analyzed. In order to validate the performance of data mining techniques, each data set is divided into 80% of training set and the remaining 20% were used for testing the model.

We evaluated the performance of models built using data mining techniques using Mean Absolute Error and Root Mean Square Error. For the purpose of model building, we considered "return" as a dependent variable. Daily closing values were transformed into daily returns using the following formula:

$$\text{Daily Returns} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Where P_t is Current closing price and P_{t-1} is Previous day closing price

2.3 Data Analysis

In this study, we used two data mining techniques namely Neural Networks and Support Vector Machines for predicting the selected indices.

Neural Networks: Neural Networks (NN) are signal processing systems or artificially created systems which are inspired by biological nervous system [44]. NN has powerful pattern classification and recognition capabilities due to their nonlinear nonparametric adaptive-learning properties. Stock market prediction is one of the major application domains of neural networks. The main advantage of neural networks is that they can estimate any nonlinear function to a random degree of accuracy with a suitable number of hidden units [45].

In current study, we employed multilayer perceptron classifier that uses backpropagation to classify instances. The levels of learning rate (lr), momentum constant (mc) and number of epochs to train through were 0.3, 0.2 and 500 respectively.

Table 1. List of selected stock indices

Selected stock index (selected stock exchange: Country)	Notation used in the study
Dow Jones Industrial Average (New York Stock Exchange: United States)	DJIA
FTSE 100 (London Stock Exchange Group: United Kingdom)	FTSE
Nikkei 225 (Japan Exchange Group-Tokyo: Japan)	NIKKEI
SSE 50 (Shanghai Stock Exchange: China)	SSE
iBovespa (BM&F Bovespa: Brazil)	IBOVESPA
Nifty 50 (National Stock Exchange: India)	NIFTY
JALSH (JSE Limited (Johannesburg): South Africa)	JALSH

Table 2. Selected technical indicators (input variables) and their description

Input variables	Description	Formula
[a] Stochastic %K	Relative position measure based on range of closing price	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$ Where Ct is closing price, LLt is lowest low and HHt is highest high in t days.
[a] Stochastic %D	Moving average of %K	$\frac{\sum_{i=0}^{n-1} \%K_{t-1}}{n}$
[a] Stochastic slow %D	Moving average of %D	$\frac{\sum_{i=0}^{n-1} \%D_{t-1}}{n}$
[b] Momentum	It measures the amount that a price has changed over a given time span	$C_t - C_{t-n}$ Where n=10, Ct is closing price today
[c] ROC (rate of change)	It measures the difference between the current price and the price n days ago	$\frac{C_t}{C_{t-n}} \times 100$
[d] LW %R (Larry William's %R)	It is a momentum indicator that measures overbought/oversold levels	$\frac{H_n - C_t}{H_n - L_n} \times 100$
[b] A/D Oscillator ((accumulation/distribution oscillator)	It is a momentum indicator that associates changes in price	$\frac{H_t - C_{t-1}}{H_t - L_t} \times 100$
[e] Disparity 5-days	It measures the relative position of the closing price to a 5-day moving average	$\frac{C_t}{MA_5} \times 100$ Where MA5 is 5-day moving average
[e] Disparity 10-days	It measures the relative position of the closing price to a 10-day moving average	$\frac{C_t}{MA_{10}} \times 100$ Where MA10 is 10-day moving average
[d] OSCP (Price Oscillator)	It displays the difference between two moving averages of a security's price	$\frac{MA_5 - MA_{10}}{MA_5}$
[a] CCI (Commodity Channel Index)	It is a measure of the deviation of the current price from the previous *n* days	$\frac{H_t + L_t + C_t - ADP_{t-1}}{0.015 \times AvgDev_{t-1}}$ Where, $ADP_t = \frac{\sum_{i=t-n+1}^{t}(H_i+L_i+C_i)}{n}$. $AvgDev_t = \frac{\sum_{i=t-n+1}^{t}\lvert H_i+L_i+C_i-ADP_{t-1})}{n}$
[a] RSI (Relative Strength Index)	It is a momentum oscillator that measures the speed and change of price movements ranges from 0 to 100	$100-\frac{100}{1+RS}$, where $RS=\frac{AU}{AD}$ *AU* = total of the upwards price changes during the past 14 days, *AD* = the total of the downwards price changes (used as positive numbers) during the past 14 days

([a] [39], [b] [40], [c] [41], [d] [42], [e] [43])

Support Vector Machines: Support Vector Machine (SVM) is a supervised statistical learning technique [46] based on Structural Risk Minimization (SRM) principle and is an approximation implementation of the method of SRM with a good generalization capability. This technique came up as a promising alternative to NN in terms of accuracy. They are less prone to overfitting than other methods. Even when the dimensionality of the data is high, SVM with a small number of support vectors can have good generalization [47]. Kernel functions play a vital role in pattern recognition through SVM.

There are various kernels for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space [34]. There are many possible kernel functions like Gaussian, Linear, Polynomial, Radial basis and Sigmoidal functions. The choice of kernel function is a critical decision for prediction efficiency. In most cases support vector machine gives better results when radial basis function (RBF) kernel is used [48]. For the current study, RBF kernel is selected for training the model.

RBF kernel:

K(x, y) = e^-(gamma * <x-y, x-y>),

where gamma is the constant of RBF.

John Platt's sequential minimal optimization algorithm was implemented using Weka software. The levels for various parameters considered in current study i.e. Complexity parameter (c), ε parameter, Tolerance parameter and Gamma of kernel function are 1, 1.0E-12, 0.001 and 0.01 respectively.

As the current study attempts to compare performance of NN and SVM across global stock indices, model specifications for a given technique are kept uniform so as to avoid any biases on account of model optimization in different settings.

2.4 Wilcoxon Signed Rank Test

Frank Wilcoxon proposed Wilcoxon signed rank test in 1945 [49]. It is a non-parametric statistical test to compare two related samples or repeated measurements on a single sample to assess whether their population mean differs [50]. It is also known as paired difference test. This test is applied to find the significant difference in hit ratio and returns across the models of all indices.

Let $\mathrm{D_i}$ be the difference between two paired random variables, assuming the difference be mutually independent, Di, i = 1, 2, … N derives from a continuous distribution F which is symmetric about a median Θ

$$\mathrm{D_i} = \mathrm{Y_i} - \mathrm{X_i},\ \text{i= 1 to N}$$

Further, $\mathrm{N_0}$ and M are denoted for number of zero and the number of non-zero differences in the sample respectively.

$$\mathrm{N} = \mathrm{N_0} + \mathrm{M}$$

Wilcoxon Signed Rank test statistic is the linear rank statistic $\mathrm{R_+} = \sum_{\mathrm{i=1}}^{\mathrm{N}} (\mathrm{R_I V_i})$ where Vi = $1_{\mathrm{Di>0}}$ is the indicator for the sign of the difference and $\mathrm{R_i}$ is the rank of | $\mathrm{D_i}$ |, i=1,2……,N. Therefore, Wilcoxon Signed Rank test statistic represents the sum of the positive signed ranks build in terms of the sum of negative signed ranks, R− or the difference of both R = (R+) – (R−). Let $\mathrm{w_\alpha}$ be critical values for the exact distribution of R+. Reject the null hypothesis at the α level of significance if R+ ≥ $\mathrm{w_{\alpha/2}}$ or R+ ≤

$$\mathrm{R_+} = 1 + \frac{\mathrm{N(N+1)}}{2} - \mathrm{w_{\alpha/2}}.$$

Large-sample approximation uses asymptotic normal distribution of R+. Under the null hypothesis,

$$\mathrm{E_0}(R_+) = \frac{\mathrm{N(N+1)}}{4}$$

$$\mathrm{Var_0}(R_+) = \frac{\mathrm{N(N+1)(2N+1)}}{4}$$

Standardized version of R_+ is asymptotically:

$$R_+^* = \frac{R_+ - \mathrm{E_0}(R_+)}{Var_0(R_+)^{1/2}} \sim \mathrm{N}\ (0,1)$$

Reject null hypotheses if $|R_+^*| \geq Z_{1-\alpha/2}$ [51]

Framework of methodology used for conducting the study is presented as Fig. 1.

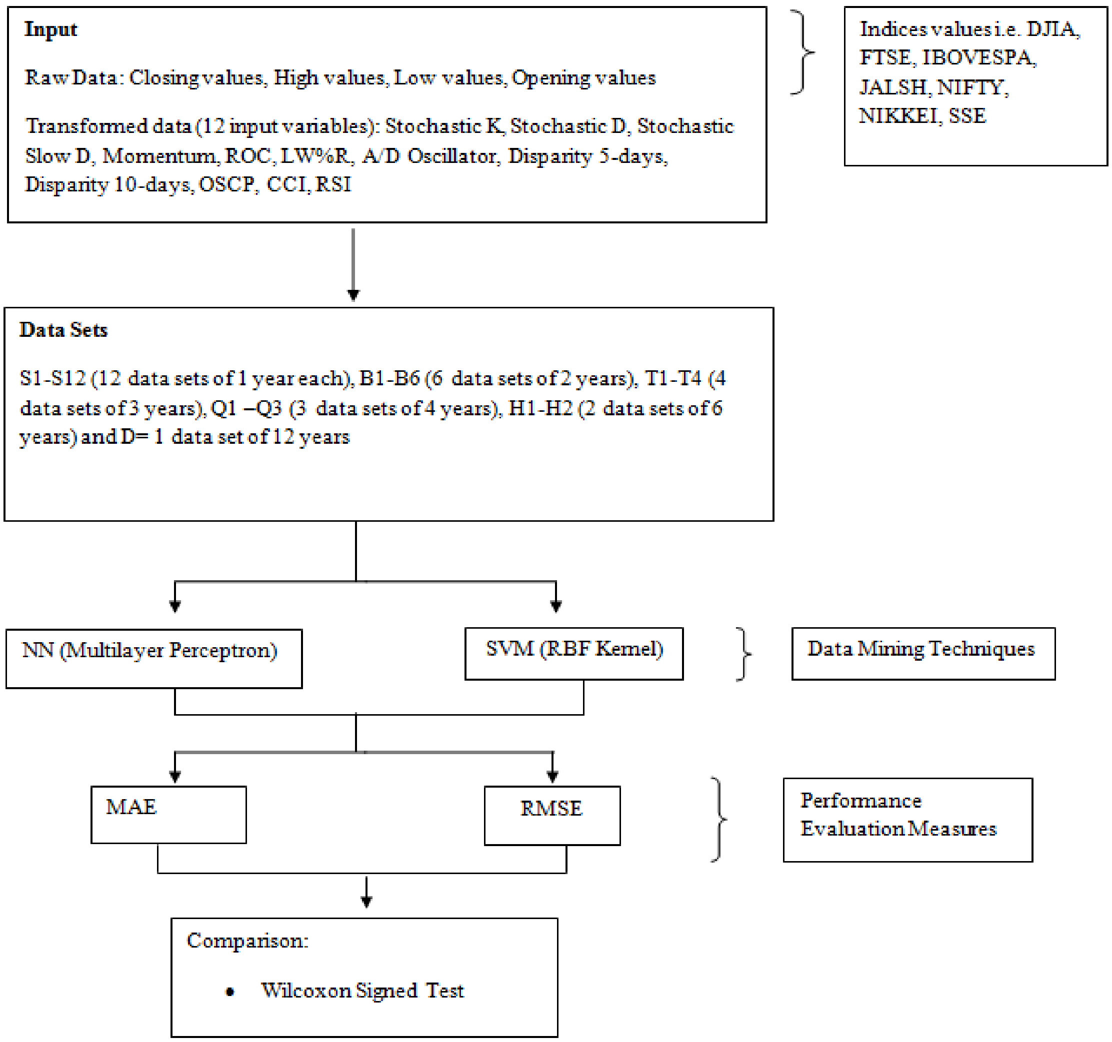

Fig. 1. Process followed under methodology

3. RESULTS

Table 3 summarizes the index wise descriptive statistics of mean absolute error for NN models of selected indices for return as a dependent variable. NN model of NIKKEI and SSE data has obtained highest mean values of mean absolute error i.e. 0.006 followed by NN model of IBOVESPA and NIFTY data with mean values of 0.005 and 0.004 respectively. NN model of FTSE and DJIA data have obtained minimum mean absolute error i.e. 0.003. Table 3 also reveals that NN models of all indices except JALSH has obtained the highest value of standard deviation i.e. 0.002. Further, NN model of NIFTY and SSE data have found to obtain the highest value of range i.e.0.009. NN model of JALSH data has obtained the minimum value of range i.e. 0.005.

Table 4 summarized index wise root mean square error for model built using NN, calculated with respect to return as an output variable. NN model of NIKKEI data has attained highest mean value of root mean square error i.e. 0.009 followed by NN model SSE and IBOVESPA data with mean values of 0.008 and 0.007 respectively. NN models of FTSE and DJIA data have observed to have minimum root mean square error i.e. 0.004. Table 4 also reveals that NN models of IBOVESPA data has obtained the highest value of standard deviation i.e. 0.005 and models of FTSE and JALSH data have obtained lowest values of standard deviation i.e.0.002. Further, NN model of IBOVESPA data has found to obtain the highest value of range i.e. 0.024. NN model of JALSH data has obtained the minimum value of range i.e. 0.008.

Table 5 summarizes the index wise descriptive statistics of mean absolute error for SVM models of selected indices. SVM models of NIKKEI and SSE have obtained highest mean values of mean absolute error i.e. 0.005 followed by SVM models of IBOVESPA and NIFTY with mean value of 0.004. SVM model of JALSH, FTSE and DJIA have observed to have minimum mean absolute error i.e. 0.003. SVM models of all indices has obtained same value of standard deviation i.e. 0.002.

Further, SVM model of DJIA has found to obtain the highest value of range i.e.0.011. SVM model of JALSH has obtained the minimum value of range i.e. 0.007.

Table 6 summarizes index wise root mean square error for model built using SVM. SVM model of SSE has attained highest mean value of root mean square error i.e. 0.008 followed by SVM model of NIKKEI with mean value 0.007 and NIFTY with mean value 0.006. SVM models of JALSH, FTSE and DJIA have observed to have minimum root mean square error i.e. 0.004. SVM models of DJIA, IBOVESPA, NIFTY, NIKKEI and SSE has obtained the highest value of standard deviation i.e. 0.004 and models of FTSE and JALSH have obtained lowest values of standard deviation i.e.0.003. Further, SVM model of IBOVESPA has found to obtain the highest value of range i.e.0.018. SVM model of JALSH has obtained the minimum value of range i.e. 0.011.

Table 3. Summary statistics of mean absolute error (NN models)

Parameter/Index	DJIA	FTSE	IBOVESPA	JALSH	NIFTY	NIKKEI	SSE
Mean	0.003	0.003	0.005	0.003	0.004	0.006	0.006
Standard Error	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.003	0.003	0.004	0.003	0.004	0.005	0.006
Standard Deviation	0.002	0.002	0.002	0.001	0.002	0.002	0.002
Sample Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kurtosis	5.339	1.945	1.626	0.977	2.700	0.231	-0.926
Skewness	2.283	1.441	1.468	1.350	1.436	1.045	0.210
Range	0.008	0.006	0.008	0.005	0.009	0.008	0.009
Minimum	0.001	0.001	0.002	0.002	0.002	0.003	0.002
Maximum	0.009	0.008	0.011	0.007	0.011	0.011	0.011

Table 4. Summary statistics of Root mean square error: NN models

Parameter/Index	DJIA	FTSE	IBOVESPA	JALSH	NIFTY	NIKKEI	SSE
Mean	0.004	0.004	0.007	0.005	0.006	0.009	0.008
Standard Error	0.001	0.000	0.001	0.000	0.001	0.001	0.001
Median	0.003	0.003	0.005	0.004	0.005	0.007	0.008
Standard Deviation	0.003	0.002	0.005	0.002	0.003	0.004	0.003
Sample Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kurtosis	6.103	4.866	8.014	2.187	1.105	-0.971	-1.261
Skewness	2.507	1.996	2.724	1.596	1.254	0.728	0.043
Range	0.012	0.011	0.024	0.008	0.012	0.013	0.011
Minimum	0.001	0.002	0.003	0.003	0.003	0.004	0.002
Maximum	0.013	0.013	0.027	0.010	0.015	0.017	0.013

Table 5. Summary statistics of mean absolute Error: SVM models

Parameter/Index	DJIA	FTSE	IBOVESPA	JALSH	NIFTY	NIKKEI	SSE
Mean	0.003	0.003	0.004	0.003	0.004	0.005	0.005
Standard Error	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.002	0.002	0.003	0.003	0.003	0.004	0.005
Standard Deviation	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Sample Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kurtosis	6.807	4.845	6.786	4.897	2.872	1.201	-0.886
Skewness	2.626	2.193	2.453	2.191	1.784	1.395	0.351
Range	0.011	0.009	0.010	0.007	0.009	0.008	0.009
Minimum	0.001	0.001	0.002	0.002	0.002	0.003	0.002
Maximum	0.012	0.011	0.012	0.009	0.011	0.011	0.010

Table 6. Summary statistics of root mean square error: SVM models

Parameter/Index	DJIA	FTSE	IBOVESPA	JALSH	NIFTY	NIKKEI	SSE
Mean	0.004	0.004	0.006	0.004	0.006	0.007	0.008
Standard Error	0.001	0.001	0.001	0.000	0.001	0.001	0.001
Median	0.003	0.003	0.004	0.004	0.004	0.007	0.007
Standard Deviation	0.004	0.003	0.004	0.003	0.004	0.004	0.004
Sample Variance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kurtosis	7.678	5.154	6.871	4.477	2.604	1.498	-1.190
Skewness	2.769	2.275	2.413	2.151	1.796	1.294	0.298
Range	0.016	0.014	0.018	0.011	0.013	0.014	0.012
Minimum	0.001	0.002	0.003	0.002	0.002	0.003	0.002
Maximum	0.018	0.016	0.021	0.013	0.016	0.018	0.014

We compared various indices in pairs so as to ascertain comparative performance of predictive models across different indices. For making the comparison of MAE and RMSE, of all possible pairs of indices, we applied Wilcoxon Signed Rank test, a non-parametric test.

The comparison of Mean Absolute Error and Root Mean Square Error of all index pairs using NN model, where dependent variable is return, is presented in Table 7. There was a significant difference between Mean Absolute Error of different indices based on NN models for 18 out of 21 index pairs. Maximum mean difference of 0.003 is observed for NIKKEI-FTSE, NIKKEI-DJIA, SSE-DJIA, SSE-FTSE and SSE-JALSH. Minimum mean difference of -0.001 is observed for JALSH-IBOVESPA. There is no significant difference between the MAE of FTSE-DJIA, NIFTY-IBOVESPA and SSE-NIKKEI. Further, that there was a significant difference between Root Mean Square Error of different indices based on NN models for 17 out of 21 index pairs. Maximum mean difference of 0.00458 is observed for NIKKEI-DJIA. Minimum mean difference of 0.00013 is observed for FTSE-DJIA.

The comparison of Mean Absolute Error and Root Mean Square Error of all index pairs using SVM model, where dependent variable is return, is presented in Table 8. There was a significant difference between Mean Absolute Error of different indices based on SVM models for 17 out of 21 index pairs. Maximum mean difference of 0.00244 is observed for SSE-DJIA and SSE-FTSE. Minimum mean difference of -0.0001 is observed for NIFTY-IBOVESPA. Further, there was a significant difference between Root Mean Square Error of different indices based on NN models for 17 out of 21 index pairs. Maximum mean difference of 0.00346 is observed for SSE-DJIA. Minimum mean difference of 0.00012 is observed for JALSH-FTSE.

Table 9 summarizes difference in accuracy measures i.e. Mean Absolute Error and Root Mean Square Error of models built using Neural Networks and Support Vector Machines. Mean absolute error as well as root mean square error were more in case of NN compared to SVM.

4. DISCUSSION

The findings of the study indicate that data mining techniques exhibits significantly different performance across selected stock indices. For NN, significant differences are observed in 18 pairs out of total 21 pairs on basis of MAE. Pair-wise comparison based on RMSE exhibit significant difference in 17 pairs out of 21 pairs of stock indices. For SVM, almost similar variation across indices is observable. On the basis of MAE, significant differences across 17 pairs are there out of total 21 pairs. On the other hand, for RMSE significant differences are there for 17 pairs out of 21 pairs. This clearly indicates a difference in performance of data mining techniques across different countries. These differences may be on account of reasons such as structure of the market, level of maturity, market stability, risk factor, volatility of market, political stability [16,52]. Findings of the study also indicate that SVM perform fair better than NN both in terms of Mean Absolute Error and Root Mean Square Error. The superiority of SVM in terms of error measures is supported by various studies [32,53,54].

Table 7. MAE and RMSE based comparison of Index pairs: NN models

Pairs	MAE		RMSE	
	Mean difference (Std error)	***S*- value (*p*- value)**	**Mean difference (Std Error)**	***S*- value (*p*- value)**
FTSE vs DJIA	0.000 (0.000)	21.5 (0.6327)	0.00013 (0.00027)	36 (0.422)
IBOVESPA vs DJIA	0.002 (0.000)	184.5 (<.0001)	0.00273 (0.00082)	194 (<.0001)
IBOVESPA vs FTSE	0.002 (0.000)	175 (<.0001)	0.0026 (0.00085)	173 (<.0001)
JALSH vs DJIA	0.001 (0.000)	96 (0.0179)	0.0006 (0.00032)	98.5 (0.0219)
JALSH vs FTSE	0.001 (0.000)	115 (0.0063)	0.00048 (0.00027)	82.5 (0.0586)
JALSH vs IBOVESPA	-0.001 (0.000)	-125.5 (0.0012)	-0.0021 (0.00089)	-139.5 (0.0006)
NIFTY vs DJIA	0.001 (0.000)	158 (<.0001)	0.00197 (0.00053)	158 (<.0001)
NIFTY vs FTSE	0.001 (0.000)	171 (<.0001)	0.00185 (0.00047)	154.5 (<.0001)
NIFTY vs IBOVESPA	0.000 (0.000)	-26.5 (0.5341)	-0.0008 (0.00091)	-19 (0.6565)
NIFTY vs JALSH	0.001 (0.000)	93 (0.0313)	0.00137 (0.00045)	111.5 (0.0084)
NIKKEI vs DJIA	0.003 (0.000)	202 (<.0001)	0.00458 (0.00074)	202 (<.0001)
NIKKEI vs FTSE	0.003 (0.000)	202 (<.0001)	0.00445 (0.00072)	201.5 (<.0001)
NIKKEI vs IBOVESPA	0.001 (0.000)	119 (0.0045)	0.00185 (0.00117)	107.5 (0.0115)
NIKKEI vs JALSH	0.002 (0.000)	194.5 (<.0001)	0.00398 (0.00074)	192 (<.0001)
NIKKEI vs NIFTY	0.001 (0.000)	135 (0.0009)	0.00261 (0.00089)	124.5 (0.0027)
SSE vs DJIA	0.003 (0.000)	189 (<.0001)	0.00385 (0.00064)	172 (<.0001)
SSE vs FTSE	0.003 (0.000)	202 (<.0001)	0.00373 (0.00055)	198.5 (<.0001)
SSE vs IBOVESPA	0.001 (0.000)	117.5 (0.0051)	0.00113 (0.00107)	88 (0.0316)
SSE vs JALSH	0.003 (0.000)	184.5 (<.0001)	0.00325 (0.00057)	172.5 (<.0001)
SSE vs NIFTY	0.002 (0.000)	145 (<.0001)	0.00188 (0.00059)	111.5 (0.0084)
SSE vs NIKKEI	0.000 (0.001)	28.5 ((0.5261)	-0.0007 (0.00098)	-20 (0.6571)

(Text in bold represents significant difference in index pairs)

Table 8. MAE and RMSE based comparison of Index pairs: SVM Models

Pairs	MAE		RMSE	
	Mean difference (std error)	***S*- value (*p*- value)**	**Mean difference (Std error)**	***S*- value (*p*- value)**
FTSE vs DJIA	-1.00E-06 (0.00017)	7.5 (0.8603)	0.00016 (0.00023)	51.5 (0.2464)
IBOVESPA vs DJIA	0.00103 (0.00022)	160.5 (<.0001)	0.00184 (0.00038)	164 (<.0001)
IBOVESPA vs FTSE	0.00104 (0.00014)	187 (<.0001)	0.00169 (0.00032)	198 (<.0001)
JALSH vs DJIA	0.00026 (0.00023)	56.5 (0.0853)	0.00028 (0.00034)	90 (0.0274)
JALSH vs FTSE	0.00027 (0.00015)	93 (0.0310)	0.00012 (0.00022)	45 (0.2872)
JALSH vs IBOVESPA	-0.0008 (0.00017)	-163.5 (<.0001)	-0.00160 (0.00041)	-168 (<.0001)
NIFTY vs DJIA	0.00098 (0.00035)	116 (0.0032)	0.00137 (0.00051)	128 (0.0020)
NIFTY vs FTSE	0.00099 (0.00022)	157.5 (<.0001)	0.00121 (0.00035)	120 (0.0010)
NIFTY vs IBOVESPA	-0.0001 (0.00028)	-23 (0.5687)	-0.00050 (0.00050)	-45 (0.2878)
NIFTY vs JALSH	0.00071 (0.00027)	86 (0.0359)	0.00109 (0.00038)	107 (0.0074)
NIKKEI vs DJIA	0.00196 (0.00028)	196 (<.0001)	0.00302 (0.00050)	179.5 (<.0001)
NIKKEI vs FTSE	0.00196 (0.00022)	201 (<.0001)	0.00286 (0.00043)	202 (<.0001)
NIKKEI vs IBOVESPA	0.00093 (0.00025)	121.5 (0.0008)	0.00118 (0.00053)	84 (0.0410)
NIKKEI vs JALSH	0.00169 (0.00022)	201 (<.0001)	0.00274 (0.00044)	186 (<.0001)
NIKKEI vs NIFTY	0.00098 (0.00028)	125 (0.0013)	0.00165 (0.00050)	126.5 (0.0023)
SSE vs DJIA	0.00244 (0.00042)	166 (<.0001)	0.00346 (0.00067)	156 (<.0001)
SSE vs FTSE	0.00244 (0.00038)	197 (<.0001)	0.00331 (0.00061)	176 (<.0001)
SSE vs IBOVESPA	0.0014 (0.00037)	131 (0.0006)	0.00162 (0.00071)	104 (0.0149)
SSE vs JALSH	0.00217 (0.00038)	181.5 (<.0001)	0.00319 (0.00060)	1820 (<.0001)
SSE vs NIFTY	0.00146 (0.00039)	129.5 (0.0017)	0.00209 (0.00065)	109 (0.0062)
SSE vs NIKKEI	0.00048 (0.00045)	24 (0.5938)	0.00045 (0.00084)	7(0.8767)

(Text in bold represents significant difference in index pairs)

Table 9. Difference in results of NN and SVM

SVM-NN (Output variable)	SVM	NN	Mean difference	Std error	Correlation	Test statistic *S* (*p*-value)
MAE (Return)	0.0039	0.0042	-0.0003	0.0001	0.886	-2690 (0.0001)
RMSE (Return)	0.0056	0.0060	-0.0004	0.0002	0.838	-1883.5 (0.0100)

5. CONCLUSION

Present study attempts to compare the performance of data mining techniques, i.e., Neural Network and Support Vector Machines across global stock indices. Results of the study are based on data mining models built using 12 input variables. Model performance is evaluated and compared on basis of Mean Absolute Error and Root Mean Square Error. Findings of the study indicate that NN and SVM exhibit significantly different accuracy across the global stock indices. Significant differences, across the global indices, in terms of MAE and RMSE, throw open an interesting research domain that needs to be explored further. Also, Mean Absolute Error and Root Mean Square Error of predicted values of daily returns for models built using NN were greater than Mean Absolute Error and Root Mean Square Error of models built using SVM. Findings of the study carry useful insights for academicians, researchers and practitioners in the domain of data mining based modeling in stock market. The study can be extended by considering different levels for different parameters of data mining techniques like number of hidden layers and nodes in Artificial Neural Networks, kernel function in support vector machines.

ACKNOWLEDGEMENT

Authors acknowledge the support of Indian Council of Social Science Research (ICSSR) for carrying out this study.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Weiss SH, Indurkhya N. Predictive data mining: A practical guide. Morgan Kaufmann Publishers, San Francisco, CA; 1998.
2. Nag BN, Han C, Yao DQ. Information enhancement in data mining: A study in data reduction. International Journal of Data Analysis Techniques and Strategies. 2015;7(1):3–20.
3. Fayyad U, Djorgovski SG, Weir N. Automating the analysis and cataloging of sky surveys. In Fayyad U, Shapiro GP, Smyth P, Uthurusamy R. (Eds). Advances in Knowledge Discovery and Data Mining. Cambridge: MIT Press. 1996;471-94.
4. Shapiro GP, Frawley WJ, Matheus CJ. Knowledge discovery in databases: An overview. AI Magazine. 1992;13:57–70.
5. Chen WH, Shih JY, Wu S. Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. Int J Electronic Fin. 2006;1:49-67.
6. Kumar DA, Ravi V. Predicting credit card customer churn in banks using data mining. Int J Data Analysis Techniques & Strategies. 2008;1:4-28.
7. Rygielski C, Wang JC, Yen DC. Data mining techniques for customer relationship management. J Tech Society. 2002;24:483–502.
8. Paramjeet, Ravi V. Bacterial foraging trained wavelet neural networks: Application to bankruptcy prediction in banks. Int J Data Analysis Techniques & Strategies. 2011;3:261–280.
9. Ramu K, Ravi V. Privacy preservation in data mining using hybrid perturbation methods: An application to bankruptcy prediction in banks. Int J Data Analysis Techniques & Strategies. 2009;1:313-31.
10. Rahman SMM, Siddiky FA, Shrestha U. A statistical data mining approach in bacteriology for bacterial identification. Int J Data Analysis Techniques & Strategies. 2011;3:117–142.
11. Lahoti G, Pratihar DK. Recurrent neural networks to model input-output relationships of metal inert gas (MIG) welding process. Int J Data Analysis Techniques & Strategies. 2017;9:248-82.
12. Yang H, Chan L, King I. Support vector machine regression for volatile stock market prediction. In Proceedings of the third International Conference on Intelligent Data Engineering and Automated Learning. 2002;391–396.
13. Muralidharan V, Sugumaran V. SVM-based wavelet selection for fault diagnosis of mono block centrifugal pump. International Journal of Data Analysis Techniques and Strategies. 2016;8:357-69.
14. Saimurugan M, Ramachandran KI. A comparative study of sound and vibration signals in detection of rotating machine faults using support vector machine and independent component analysis. Int J Data Analysis Techniques & Strategies. 2014;6:188-204.

15. Erturk E, Sezer EA. Software fault prediction using Mamdani type fuzzy inference system. Int J Data Analysis Techniques & Strategies. 2016;8:14-28.
16. Ahmad N, Tanaka T, Saito Y. Machining parameter optimisation by genetic algorithm and artificial neural network. Int J Data Analysis Techniques & Strategies. 2014;6:261-274.
17. Tigas G, Lefakis P, Ioannou K, Hasekioglou A. Evaluation of artificial neural networks as a model for forecasting consumption of wood products. Int J Data Analysis Techniques & Strategies. 2013;5:38–48.
18. Tuckova J, Sramka M. ANN application in emotional speech analysis. Int J Data Analysis Techniques & Strategies. 2012;4:256–276.
19. Taylor Q, Giraud-Carrier C, Knutson CD. Applications of data mining in software engineering. Int J Data Analysis Techniques and Strategies. 2010;2:243–257.
20. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. J Computational Sci. 2011;2:1–8.
21. Kuo RJ, Chen CH, Hwang YC. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. Fuzzy Sets and Systems. 2001;118:21-45.
22. Michael G, Connor O, Madden MG, Connor NO, Madden MG. A neural network approach to predicting stock exchange movements using external factors. Proceedings of AI-2005, 25th International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge; 2005.
23. Mittal A, Goel A. Stock prediction using twitter sentiment analysis. Stanford University Working Paper; 2012.
24. Thawornwong S, Enke D. The adaptive selection of financial and economic variables for use with artificial neural networks. Neurocomputing. 2004;56:205–232.
25. Mizuno H, Kosaka M, Yajima H. Application of neural network to technical analysis of stock market prediction. Studies in Informatics and Control. 1998;7:111-120.
26. Majumder M, Hussian MD. Forecasting of Indian stock market index using artificial neural network; 2010.
Available:https://www.nse-india.com/content/research/FinalPaper206.pdf
27. Vojinovic Z, Kecman V, Seidel R. Modelling empirical data and decision making with neural networks to financial time series. International Journal of Management and Decision Making. 2002;3(2).
28. Tjung LC, Kwon O, Tseng KC, Geist AB. Forecasting financial stocks using data mining. J Global Eco Fin. 2010;3:13-26.
29. Hammad AAA, Ali SMA, Hall EL. Forecasting the jordanian stock prices using artificial neural network; 2009.
Available:http://www.min.uc.edu/robotics/papers/paper2007/Final%20NNIE%2007%20SOUMA%20Alhaj%20Ali%206p.pdf)
30. Altay E, Satman MH. Stock market forecasting : Artificial neural network and linear regression comparison in an emerging market. Journal of Financial Management and Analysis. 2005;18(2):18-33.
31. Cao L, Tay FEH. Financial forecasting using support vector machines. Neural Comput & Applic. 2001;10:184-92.
32. Kim K. Financial time series forecasting using support vector machines. Neurocomputing. 2003;55:307-19.
33. Huang W, Nakamori Y, Wang S. Forecasting stock market movement direction with support vector machine. Computer and Operation Research. 2005;32:2513-22.
34. Kumar M, Thenmozhi M. Forecasting stock index movement: A comparison of support vector machines and random forest. Indian Institute of Capital Markets 9th Capital Markets Conference Paper, 2005; 2006.
Available:http://ssrn.com/abstract=876544
35. Anonymous; 2016.
[Retrieved on 18th March]
Available:https://www.msci.com/market-classification
36. Liu H, Setino R. A probabilistic approach to feature selection. In: ML Proceedings 13th ICML, 1996;319-327.
37. Asadi S, Hadavandi E, Mehmanpazir F, Masoud M, Nakhostin MM. Hybridization of evolutionary Levenberg- Marquardt neural networks and data pre- processing for stock Market prediction. Knowledge-Based Systems. 2012;35:245–258.
38. Kim K. Artificial neural networks with evolutionary instance selection for financial

forecasting. Expert Sys Apps. 2006;30:519-26.
39. Kaufman PJ. Trading systems and methods. New York, NY: John Wiley & Sons; 2013.
40. Chang JY, Jung K, Yeon J, Jun D, Shin, Kim H. Technical indicators and analysis methods. Jinritamgu Publishing, Seoul; 1996.
41. Murphy JJ. Technical analysis of the futures markets: A comprehensive guide to trading methods and applications. Prentice-Hall, New York; 1986.
42. Achelis SB. Technical analysis from A to Z. Chicago: Probus Publishing, Chicago; 1995.
43. Choi J. Technical indicators. Jinritamgu Publishing, Seoul; 1995.
44. Preethi G, Santhi B. Stock market forecasting techniques: A survey. Journal of Theoretical and Applied Information Technology. 2012;46:24-30.
45. Kim HJ, Shih KS. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. Applied Soft Computing. 2007;7(2):569-576.
46. Vapnik V. Statistical learning theory. Wiley, New York; 1998.
47. Han J, Kamber M, Pei J. Data mining concepts and techniques, Third edition, Elsevier Inc., Waltham, USA; 2012.
48. Arasu BS, Jeevananthan M, Thamaraiselvan N, Janrthanan B. Performances of data mining techniques in forecasting stock index – evidence from India and US. J Natn Sci Foundation. 2013;42:177–91.
49. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin. 1945;1:80–83.
50. Rosner B, Glynn RJ, Lee ML. The wilcoxon signed rank test for paired comparisons of clustered data. Biometrics. 2006;62:185–192.
51. Rey D, Neuhauser M. Wilcoxon-signed-rank test, in: M. Lovric (Ed.), International Encyclopedia of Statistical Science, Springer Berlin Heidelberg. 2014;1658–1659.
52. Flannery MJ, Protopapadakis AA. Macroeconomic factors do influence aggregate stock returns. The Review of Financial Studies. 2002;15:751–82.
53. Sheta A, Ahmad SEM, Faris H. A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. International Journal of Advanced Research in Artificial Intelligence. 2015;4:55-63.
DOI: 10.14569/IJARAI.2015.040710
54. Grosan C, Abraham A, Ramos V, Han SY. Stock market prediction using multi expression programming. In Proceedings of Portuguese conference of artificial intelligence, workshop on artificial life and evolutionary algorithms. Portuguese: IEEE Press. 2005;73–78.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle3.com/review-history/51284