



Machine Learning Algorithms for Time Series in Financial Markets

Mohammad Ghasemzadeh^{a*}, Naeimeh Mohammad-Karimi^a, Habib Ansari-Samani^b

^a Computer Engineering Department, Yazd University, Yazd, Iran

^b Management and Economics Department, Faculty of Economics, Yazd University, Yazd, Iran

ARTICLE INFO

Article history:

Received 01 March 2020

Accepted 03 July 2020

Keywords:

Financial markets

Stock market

Machine Learning

Forecasting

Time series

ABSTRACT

This research is related to the usefulness of different machine learning methods in forecasting time series on financial markets. The main issue in this field is that economic managers and scientific society are still longing for more accurate forecasting algorithms. Fulfilling this request leads to an increase in forecasting quality and, therefore, more profitability and efficiency. In this paper, while we introduce the most efficient features, we will show how valuable results could be achieved by the use of a financial time series technical variables that exist on the Tehran stock market. The suggested method benefits from regression-based machine learning algorithms with a focus on selecting the leading features to find the best technical variables of the inputs. The mentioned procedures were implemented using machine learning tools using the Python language. The dataset used in this paper was the stock information of two companies from the Tehran Stock Exchange, regarding 2008 to 2018 financial activities. Experimental results show that the selected technical features by the leading methods could find the best and most efficient values for the parameters of the algorithms. The use of those values results in forecasting with a minimum error rate for stock data.

1 Introduction

Forecast of financial time series is of the most critical issues in making financial decisions. In this regard, the Tehran Stock Exchange is of great importance for domestic and international financial markets [1,2]. Based on the economic events and data of the past, it provides a profitable method for the future. Financial time series forecasting is a challenging issue in the time-series field and has attracted many researcher's attention [3]. Nwwaaay,, it is eee ff tee ficccill mrrktt's maaagr'' ooncerns that individuals with different tastes selection and amounts of every kind of asset to be able to enter those markets, to recognize suitable opportunities and to gain good profit based on correct assessment. Today's world is that of change, and it is an essential factor in organizational success and survival to know what we expect in the future. Governments, shareholders and organizations managers need to forecast exports and imports, recognizing the market situation, and the organization's future, respectively [4]. Managers take personal and professional decisions based on forecasting situations in the future. In many cases, forecasting is based on present and past. They, in fact, try to link two or more valuables so that they can be used for forecasting. On the subject of forecasting at financial markets, it is a priority to analyze data [5]. Two standard methods are used for analyzing financial series: technical analysis, basic analysis.

Basic analysis is an analyzing method that works on key figures and information on companies' finan-

* Corresponding author. Tel.: +98-913-156-7972; fax: +98-35-3123-2357.
E-mail address: m.ghasemzadeh@yazd.ac.ir

cial statements, the country's macro-economy figures, and influencing agents on different economies. Therefore, selling or buying shares or any kind of asset is done after the information about the aforementioned items has been assessed [6]. Technical analysis is a method for forecasting markets based on assessing and studying price history and turnover on markets. Technical analysis is of mathematical formulas related to prices and turnover data that are used for modelling some aspects of shares prices for indices [7]. Technical analysis includes studying and assessing different indices, charts, and patterns that illustrate market trends and different shares status for investors. Machine learning methods, inspired by pattern-ology and computational learning theory, investigate the studying and constructing algorithms that are able to learn and forecast based on data. Such algorithms do not follow a gg,mmi ittt rttt sss yet, they forecast or make decisions based on modelling and sample input data. Machine learning methods are used on computational works where designing and programming exclusive algorithms with suitable function are hard or impossible.

Machine learning methods have a close relationship with computational statistics, and they often overlap with each other. This branch is about forecasting by computers, and has a secure connection with mathematical optimisation that it, in turn, introduces to the system method, theories and functions. Machine learning sometimes merges with data analysing. This sub-branch focuses on exploratory analysis of data, and it is known as non-supervised learning. A machine learning method, in the data analysis field, is a method for designing complex models and algorithms used for forecasting. In industry, it is known as predictive analysis. Data and past samples are the first steps to learn the machine. In the field of computer science, theories are put forward in that basics or extracted to learning data, so that software can be developed meeting users' needs. Therefore, software development will be made possible on learning machines so that parts of software system basics that are not extractable through analysing and designing by analyses and designers can be achieved from past data. In learning the scope, this body of extracted basics is called a model. So, the goal is that for the extracted model to be as common as possible so that it can be used with a high degree of precision for data that is not created. Financial markets play an essential role in organizing modern society, socially and economically [12]. Most assets are undoubtedly exchanged through the stock market worldwide, nowadays. National economies are profoundly affected by the stocks' value forecasting function on the stocks market. One of the most critical pieces of information for investors on stock markets is shares r r iee iffrr mtt i... Trrr ffr r ,, ttt ll y is aaarss r r iee iffrr mtt inn aaallggggigg ttt ll oo it is ivvsstr'' favorite.

However, investing in the stock market is an essential part of the economy. Therefore, forecasting, especially in developing countries like Iran, is very important, to manage the stock market for achieving stable development. That eases decision-making for the stock market executives under the current uncertainties. Investors are also able to forecast shares price or overall index and make logical decisions, accordingly. Considering the importance of the subject, machine-learning algorithms have already presented remarkable functions based on studies. Thus, most papers in shares price field center around intelligent methods, nowadays [4]. Therefore, doing more research in this field seems necessary. Since machine-learning methods are empowered to model complicated engineering problems and nonlinear systems, they are known as a suitable method for forecasting share prices. The main objective of tii s eeee ((b) EEHaeææeo λλλ ff fr csstigg Ir'' s tkkk mrrktt aaare r r iees ss a financial time series using learning algorithms. To achieve that goal, it is vital to select the most suitable f faatr rsm mī r rrr rr tt rr'' eettiggs, and effective ones in chosen algorithms.

This paper illustrates how suitable precision, as well as having access to the most effective features

can be achieved through utilising technical analysis of financial time series available on stock market data. Although technical analysis to achieve technical features has been taken advantage of in past studies and researches, it is an innovation to consider more technical features, diversity and arrangement of features of these kinds together, in this paper. The presented Innovations in this paper are as follows:

- 1- Selective data are the ones that are assessed with respect to the presented features for the first time.
- 2- Also employing the most usable learning algorithms based on regression to forecast different share prices on Tehran's stock market.
- 3- It is selecting 23 proposed features that have been calculated based on technical analysis.

2 Literature Review

Forecasting financial time series can be one of the main challenges in time series and machine learning scope. In past decades, several methods for forecasting financial markets and presenting decision-making back systems have been proposed. Soft calculation such as expert, phase, and neural systems has been used for financial series and modelling with relative success. Compared to traditional statistical forecasting methods; soft calculation techniques impose a non-linear relationship on input data distribution without having any prior knowledge. Artificial neural networks have recently become popular for forecasting financial markets. Artificial neural networks are data-based and self-minimizing the error function to recognize the time series' linear and non-linear relationships without any statistical hypothesis. For instance, Haw et al. [8] concluded that artificial neural networks work better statistical methods such as linear regression and Box Jenkins methods. A similar study was done by Pang et al. [9] show that artificial neural networks can be successfully used for modelling and forecasting nonlinear time series.

Some proposed neural networks that are widely used for forecasting financial markets are: the single-layer perceptron, multilayer perceptron. Although multilayer neural networks of more complex, they are used for modelling more than two other methods. In most researches, the MLP has been employed for learning relationship between some technical features and forecasting minimum and maximum of daily share prices. For example [10], MLP algorithm has been used to forecast Bangladesh share prices using a combination of the following: technical features of convergent-divergent mobile mean and relative power index. To categorize companies based on financial problems, the MLP algorithm has also been used for learning some financial amounts and company's balance sheet conformed with basic analysis. Cavalcant et al. [11] suggest different learning algorithms in financial markets for the testing field. Each of those algorithms performs forecasting depending on the financial time series that input valuables receive. Suggestive approaches on amounts and types of variables used for modelling monetary markets are different in different papers [13]. For instance, Chen and Haw [14] suggest ways to measure volume level and acceleration rates for forecasting testing learning algorithms to technical analysis of the body of technical features such as mobile mean, mobile illustration mean movement mean of convergence and divergence, volume ratio and relative power index. Barak and Modares [15] suggest income price ratio and cash flow through fundamental technical analysis on some economic and key features of a company such as its size.

3 Methodology

Considering the importance of the subject that was stated in the introduction and that forecasting shares prices are one of the most important subjects of financial markets (since shares prices data is of high variability, complexity, dynamics and chaos, the unknown relationship between shares prices and perfecting agents is dynamic. Therefore, the problem is how to improve forecasting precision as the financial time series using learning algorithms. To achieve that goal, it is significant to select the most suitable features and main and most effective parameter for selected algorithms.

3.1 The Proposed Method

This paper illustrates how promising precision, as well as the most effective features can be achieved using technical analysis of financial time series available on the stock market. Although technical analysis has been used to achieve technical features in past researches, considering more technical features, diversity, and arrangement of this type of features together are an innovation in this paper. Using backward and forward methods in this paper is due to considering all possible subsets of features. In this method, the best options with the least errors are selected using all possible options and giving scores to features. The proposed solution is dependent on employing machine learning algorithms based on regression with a focus on forwarding feature Selection methods, to finding the best input technical variables. The research flowchart is as follows:

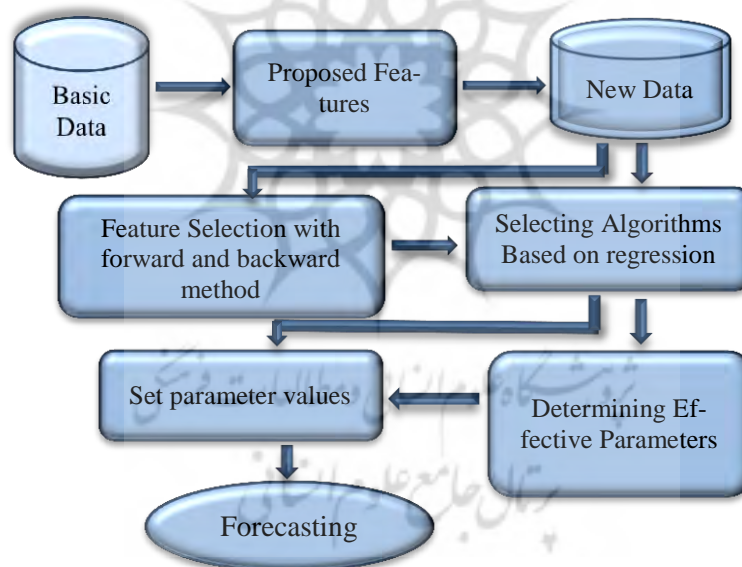


Fig. 1 The proposed Method Flowchart

The primary data is the stock information for two ... in Tehran's market. This piece of the stocks data includes ten features as follows: date, first price, highest price, lowest price, final price, volume, value, number of deals, and the share price of the last day of every day. Technical analysis was calculated using technical analysis on basic features for companies shares in a 10-year interval. New data is created with extracting technical features and adding to basic data with basic features. Proposed features that are based on technical analysis can be seen in Table 1.

Table 1: The proposed Features

Suggested feature	Num	Suggested feature	Num	Suggested feature	Num
Relative power index(RSI)	17	Moving Average Convergence-Divergence(MACD)	9	Total price index	1
Bollinger Up Band(BUB)	18	3 day moving average(3MA)	10	Industry index	2
Bollinger Lower Band(BLB)	19	5 day moving average(5MA)	11	Weight index	3
The first days of each week	20	10 day moving average(10MA)	12	Industry Indicator	4
Latest Stock Trading Days	21	20 day moving average(20MA)	13	Total Return	5
The first months of each year	22	30 day moving average(30MA)	14	Industry returns	6
Change rate index(ROC)	23	7day average of volume(7AV)	15	Beta coefficient of industry indicator	7
*Stock Returns(label Feature)	24	Weighted moving average(WMA)	16	Beta coefficient of total return	8

Table 2: Effective parameters in each algorithm

Algorithm	parameter	Values
SVR	Kernel	rbf-linear
	Gamma	0.01 ,1 ,0.001 0.1, 0.0001,
	C	10 ,15 ,20 , 100 ,130 ,136 1 ,5 ,
	epsilon	0.9 ,0.009 ,0.0009 , 0.1 ,0.001
MLP	alpha	0.01 ,0.001, 0.0001,
	hidden_layer_size	32 ,25 ,50 ,60 1 ,5 ,10 ,
	learning_rate_init	0.01 ,1 ,0.001 0.1,
	max_iter	50 ,100 ,150 ,200
	random_state	0 ,1
	tol	0.01 , 0.001,0.009 0.0001,

Three regression-based mostly used algorithms in financial markets for more scope that was chosen after that the proposed features have been collected are as follows: Support Vector Machine (SVR),

Multi-Layer Perceptron (MLP), and Decision Tree (DT). Three chosen algorithms are representatives of learning algorithm families. For example, DT algorithm, MLP algorithm, and SVR algorithm algorithm are of a tree, neural networks, and support vector family, respectively. In other words, all three algorithms are of standard and basic machine learning algorithms. Technical features are introduced into algorithms as input variables during 23 stages using forward and backward feature selection methods.

Finally, the parameters of each algorithm were investigated and those that their change of amounts was effective for forecasting in financial markets scope were selected. (Table 2) A list of amounts designated to each of those parameters during the implementation stage. The routine is for algorithms to set different amounts in each parameter list during each try and to keep the best and most suitable amounts that improve forecasting. Finally, the best amount of selected parameters of each algorithm for each input feature was selected.

4 Evaluation and Results

What is important in financial series forecasting using the learning machine algorithm is to provide the best forecasting by comparing and evaluating of different algorithms results. One of the active factors in algorithms' success is to evaluate them according to suitable criteria.

4.1 Evaluation Data

The data used in this paper is of two companies in Tehran's stock market. The data has been collected from 2008 to 2018 using the software Tesclient. The data include 2000 samples and 10 basic features: date, first price, highest bid, lowest bid, final price, volume, value, number of transactions, and average price's price. Data is collected through the first time. Then, data with no amount is removed using noise removing functions, and the data is subsequently normalized. According to the researches done on financial series forecasting using machine learning algorithms, evaluation figures are put into two categories using validation methods: train set and test set, the train set is used for training algorithms. It shall be noticed that the train set, in turn, is divided into 2 categories, train, and validation that are used for evaluating training amounts resulted from train figures. Results from the evaluation are only used for selecting the best train set. It is certain that results from the evaluation of figures used in training or not considered as general evaluation results. Yet, what is meant as a result of algorithm precision evaluation is the precision of that algorithm on forecasting samples that are in the test set category. To determine test and train set for each share, 60% of data to the training set, and 20% of data to test set is designated, respectively. The other 20% is considered as a validation set, all of which are shown in Table 3.

Table 3: The number of train and test samples for each of the two stocks

STOCK NAME	TOTAL SAMPLE	TRAIN SET	TEST SET
PARS OIL COMPANY	616	414	140
SHAZAND PETROCHEMICAL COMPANY.	716	443	138

In the financial time series, traditional mutual evaluation (like K-fold) cannot be used due to their time dependence and arbitrary selecting of the test set. Therefore, validation stages used in financial time series are as follows:

First stage: all data is divided into some parts. The assumption for the implementation is three, but the number 10 was considered as a suitable number for dividing the data. Therefore, the data is divided into 10 parts and one part is chosen during each stage.

The second stage: The selected part of 10 parts is divided into test and train sets, and then the validation is done.

The third stage: again, another part of 10 selected parts and the part that was validated during the second stage is added to this part. Then, train and test sets are specified, and evaluated. As in figure 2, the trend continues until all the 10 parts are validated.

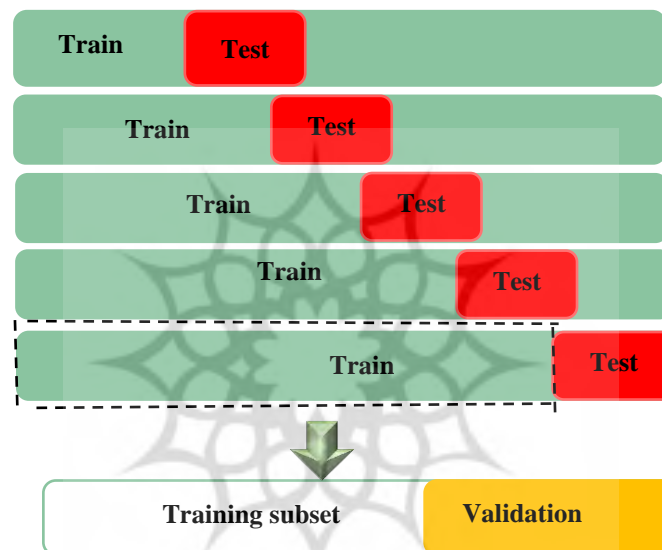


Fig. 2 Validate Financial Data Series

To accomplish the evaluations, basic figures are once evaluated with 10 basic features. Then an evaluation is done on new figures. The new figures are the ones with proposed features that are selected using forward and backward methods and are introduced into algorithms as inputs.

4.2 Evaluation Criteria

To evaluate learning algorithms using the proposed method, the standard criteria in financial series forecasting scope mean square error and root mean square error is employed. These criteria are widely used for forecasting.

1. The square mean error, as it is called, three operations are done on data:

- 1) Calculating algorithms output error,
- 2) Multiplying the error by the power of 2,
- 3) Calculating the mean of the sum of all errors that have been multiplied by the power of 2 and it can be calculated by relation 1.

$$RMSE \cong \frac{1}{N} \sum (f_i - y_i)^2 \quad (1)$$

The amount of each variable in relation 1 is:

f_i : Equivalent of algorithms output

y_i : Equivalent of a definitive answer

2. Mean square root error, which is a very useful criterion for measuring precision for continuous time that is as in relation 2.

$$RMSE \cong \sqrt{\frac{1}{N} \sum (f_i - y_i)^2} \quad (2)$$

As illustrated in Table 4, validating and testing for Pars oil stocks, on basic figures (data) show that the SVR algorithm has a better function than MLP and DT with both MSE and RMSE evaluation criteria.

Table 4: Pars oil stocks' performance results on Basic Data

Algorithm	RMSE(Validation)	SE(Validation)
SVR	<u>0.05776</u>	<u>0.00382</u>
MLP	0.05821	0.00387
DT	0.05787	0.00384
Algorithm	RMSE(Test)	MSE(Test)
SVR	0.23161	0.05364
MLP	<u>0.19640</u>	<u>0.03857</u>
DT	0.20340	0.04137

Considering results in Table 5 for those shares, all three algorithms with proposed features with respect to basic figures did the forecasting with a lower error rate. All three algorithms with proposed features have a better function than forecasting algorithms on basic data, while validating and testing. Results showed that MLP and SVR algorithms, having similar outputs, have a better function than a decision tree. However MLP with proposed features (1 of 4 cases of proposed features from the fourth stage of selecting the feature forward) total index, total efficiency, industry index, RSI index and SVR algorithm with proposed features (1 of 5 cases of proposed features from the fifth stage of selecting the feature forward) total index, total efficiency, industry index, industry efficiency, industry beta co-efficiency, and 30-day moving men have a better function than decision tree algorithm. As illustrated in Table 5, the MLP algorithm has a better function than two other algorithms for testing samples. The evaluation results showed that although four and five cases did not result in lowest error rate for testing samples with proposed features, the lowest error rate as the best forecasting can be achieved for dual cases from proposed features. Evaluation indices in the first or second stage with one feature or two features results from the lowest rate for test samples. As it is shown in Tables 5 and 6, validation forecasting and testing of three algorithms for Shazand petrochemical shares for basic data with proposed features for basic data with 10 basic features have shown improvement. Invalidation for data with

Table 5: Evaluation Results for Pars Oil Company with Propose Features

Selected Feature	Algorithm	RMSE(Validation)	MSE(Validation)
Total Return,Industry Indicator,Industry Return, Beta Industry Indicator,MA30	<u>SVR</u>	<u>0.03490</u>	<u>0.00137</u>
Total price index ,Industry Indicator,RSI	<u>MLP</u>	<u>0.03481</u>	<u>0.00135</u>
Total price index ,Total Return,Industry Indicator,Industry Return, Beta Industry Indicator,Month of year	DT	0.03898	0.00165
Selected Feature	Algorithm	RMSE(Test)	MSE(Test)
Industry Indicator,ROC	SVR	0.22742	0.05172
Latest Stock Trading Days Total price index	<u>MLP</u>	<u>0.18047</u>	<u>0.03257</u>
Total Return Beta Total Return	DT	0.19830	0.03932

Table 6: Results of the Evaluation of Shazand Petrochemical Shared with Basic Data

Algorithm	RMSE(Validation)	MSE(Validation)
SVR	0.06646	0.00580
MLP	<u>0.06532</u>	<u>0.00478</u>
DT	0.07410	0.00599
Algorithm	RMSE(Test)	MSE(Test)
SVR	0.45241	0.20467
MLP	0.42556	0.18110
DT	<u>0.41418</u>	<u>0.17155</u>

Table 7: Results of the Evaluation of Shazand Petrochemical with Propose Features

Selected Feature	Algorithm	RMSE(Validation)	MSE(Validation)
Total price index ,Total Return,Industry Indicator,Industry Return, Beta Industry Indicator,MACD	<u>SVR</u>	<u>0.02982</u>	<u>0.00101</u>
Total price index ,Total Return,Industry Indicator,Industry Return, Beta Industry Indicator,MACD	MLP	0.04086	0.00188
Total price index ,Total Return,Industry Indicator,Industry Return, Beta Total Return,3MA,BLB	DT	0.04266	0.00204
Total Return, ROC	SVR	0.40723	0.16583
Total Return, WMA	<u>MLP</u>	<u>0.32577</u>	<u>0.10612</u>
RSI	DT	0.40814	0.16658

5 Conclusions

Evaluation results on two shares with forwarding, and backward feature selecting methods for 23 proposed features based on technical analysis showed that MLP and SVR algorithms have a better function than a decision tree. Another impressive result that can be seen in evaluation is that all the stages from 1 to 23 of selecting features for validation have a lower error with respect to elementary data that the best forecasting with the lowest rate for stages 1 to 6 is seen using forward feature selecting. Among those stages, foursome or five some or six some cases have a better result than other cases. In the experiment stage, although the lowest rate was not in foursome or five-some or six-some of the proposed feature, it was in stages 1 to 6; therefore, testing results are acceptable. As can be seen in figures 3 and 4, validation results clearly show the effective presence of proposed features (total index, total efficiency, and industry index and industry co-efficiency for both MLP and SVR algorithms in different combinations from selected features in each of those two shares.

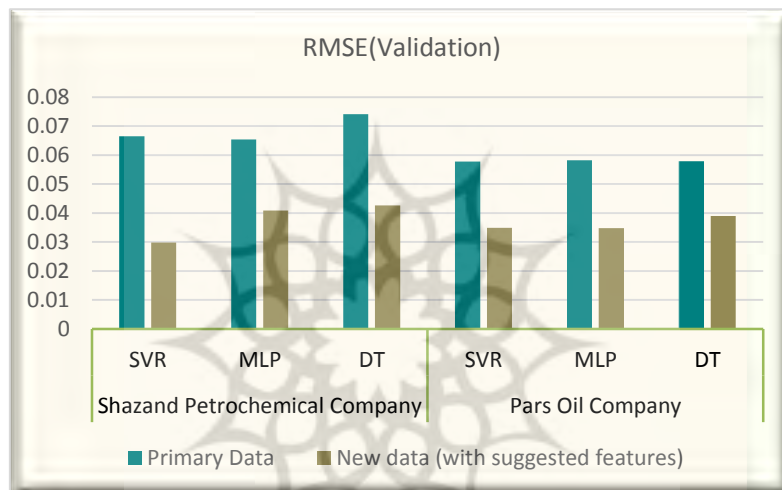


Fig.i3 Validation results with RMSE

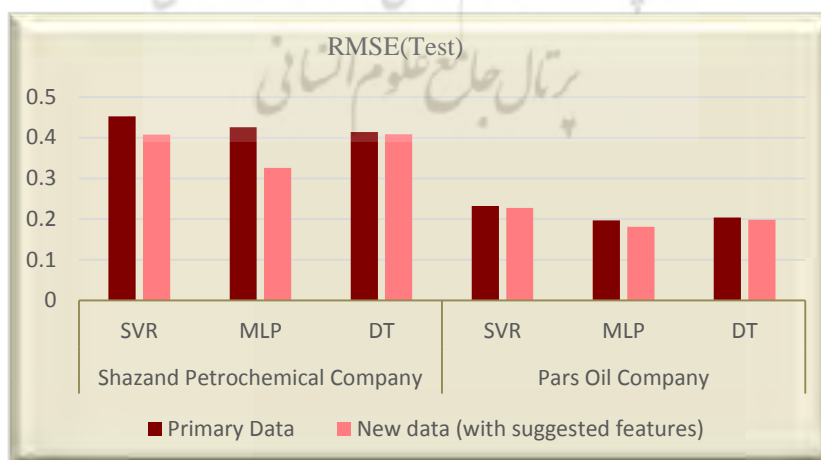


Fig. 4 Testing results with RMSE

References

- [1] Abbasi, E., Samavi, M.E., Koosha, E., *Performance Evaluation of the Technical Analysis Indicators in Comparison with the Buy and Hold Strategy in Tehran Stock Exchange Indices*, Advances in Mathematical Finance and Applications, 2020, **5**(3), P. 285-301. Doi: 10.22034/AMFA.2020.1893194.1376
- [2] Khaleghi Kasbi, P., Aghaei, M. A., Rezaei, F., *Saliency Theory and Pricing Stock of Corporates in Tehran Stock Exchange*, Advances in Mathematical Finance and Applications, 2018, **3**(4), P.1-16. Doi: 10.22034/AMFA.2018.577140.1120
- [3] Gupta, R., Pierdzioch, C., Selmi, R., Wohar, M. E., *Does partisan conflict predict a reduction in US stock market (realized) volatility? Evidence from a quantile-on-quantile regression model*, North American Journal of Economics and Finance, 2018, **43**(2), P. 87–96. Doi: 10.1016/j.najef.2017.10.006
- [4] Nadkarni, J., Ferreira Neves, R., *Combining NeuroEvolution and Principal Component Analysis to trade in the financial markets*, Expert Systems with Applications, 2018, **103**(1), P.184–195. Doi: 10.14419/ijet.v7i4.21723
- [5] Zhong, X., Enke, D., *Forecasting daily stock market return using dimensionality reduction*, Expert Systems with Applications, 2017, **67**(1), P. 126–139. Doi: 10.1016/j.eswa.2016.09.027
- [6] Ballings, M., Van Den Poel, D., Hespeels, N., Gryp, R., *Evaluating multiple classifiers for stock price direction prediction*, Expert Systems with Applications, 2015, **42**(20), P. 7046–7056. Doi: 10.1016/j.eswa.2015.05.013
- [7] Patel, J., Shah, S., Thakkar, P., Kotecha, K., *Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques*, Expert Systems with Applications, 2015, **42**(1), P. 259–268. Doi: 10.1016/j.eswa.2014.07.040
- [8] Ghiassi, M., Burnley, C., *Measuring effectiveness of a dynamic artificial neural network algorithm for classification problems*, Expert Systems with Applications, 2010, **37**(4), P. 3118–3128. Doi: 10.1016/j.eswa.2009.09.017
- [9] Pang, X., Zhou, Y., Wang, P., Lin, W., Chang, V., *An innovative neural network approach for stock market prediction*, The Journal of Supercomputing, 2018. Doi: 10.1007/s11227-017-2228-y
- [10] Hsu, M. W., Lessmann, S., Ma, T., Johnson, J. E. V., *Bridging the divide in financial market forecasting: machine learners vs. financial economists*, Expert Systems with Applications, 2016, **61**(1), P. 215–234. Doi : 10.1016/j.eswa.2016.05.033
- [11] Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., Nobrega, J. P., Oliveira, A. L. I., *Computational Intelligence and Financial Markets: A Survey and Future Directions*, Expert Systems with Applications, 2016, **55**(1), P. 194–211. Doi: 10.1016/j.eswa.2016.02.006
- [12] Izadikhah, M., *Improving the Banks Shareholder Long Term Values by Using Data Envelopment Analysis Model*, Advances in Mathematical Finance and Applications, 2018, **3**(2), P. 27-41. Doi: 10.22034/AMFA.2018.540829

[13] Gurav, U., Sidnal, N., *Predict Stock Market Behavior: Role of Machine Learning Algorithms*, in *Intelligent Computing and Information and Communication*, Springer, 2018, P. 383–394.

Doi: 10.1007/978-981-10-7245-1_38

[14] Chen, Y., Hao, Y., *A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction*, *Expert Systems with Applications*, 2017, **80**(1), P. 340–355.

Doi: 10.1016/j.eswa.2017.02.044

[15] Barak, S., Modarres, M., *Developing an approach to evaluate stocks by forecasting effective features with data mining methods*, *Expert Systems with Applications*, 2015, **42**(3), P. 1325–1339.

Doi: 10.1016/j.eswa.2014.09.026

