Journal of Financial Management Strategy
Alzahra University
Vol. 10, No. 38, Fall 2022

pp. 1-4

Research Paper

Predicting the Tehran Stock Exchange Index Using Support Vector Regression; Based on the Dimension Reduction Technique[1]

Somayeh Mohebi[2], Mohamad Esmail Fadaienejad[3], Mohamad Osoolian[4]

Received: 2021/03/30 Accepted: 2022/09/10

1. INTRODUCTION

The stock market prediction has always been challenging for researchers, analysts, and investors because it is influenced by various factors, many of which can be used as input variables in stock market prediction. Soft computing techniques have demonstrated significant prediction potential in previous research. The Support Vector Regression (SVR) technique is one of the most important techniques that has recently received much attention within financial markets. It can detect nonlinear time series behaviors without relying on statistical hypotheses and also reduce empirical risks while minimizing structural risks (Cavalcant et al., 2016). Nevertheless, if an intelligent model, such as the SVR is expected to produce an efficient and accurate prediction, it is necessary to choose the most influential and representative inputs. This type of selection is the main task of dimensionality reduction technology (Zhong and Enke, 2017).

Dimension reduction can be accomplished in two ways: 1) by selecting the most relevant variables from the main data set called feature selection, and 2) by generating a smaller group of new variables, each of which is a unique combination of older input variables, called feature extraction.

The majority of data processing research has employed some form of dimension reduction technique (feature selection or extraction). However, each of these methods

1. DOI: 10.22051/JFM.2022.35543.2528
2. Ph.D. Student, Department of Financial Management and Insurance, Faculty of Management and Accounting, Shahid Beheshti University, Email: so_mohebi@sbu.ac.ir.
3. Associate Professor, Department of Financial Management and Insurance, Faculty of Management & Accounting, Shahid Beheshti University. Corresponding Author. Email: mfadaei@sbu.ac.ir.
4. Associate Professor, Department of Financial Management and Insurance, Faculty of Management & Accounting, Shahid Beheshti University. Email: m_osoolian@sbu.ac.ir.

makes unique assumptions about selecting the most appropriate features. Feature extraction methods attempt to reduce the dimensions of the main feature set to the feature space without altering the nature of the problem. However, because the principal meaning of the features is altered during this transfer, there is no way to prioritize or identify the selected features. On the other hand, feature selection retains the original meaning of features after reduction, in contrast to feature extraction methods. However, the majority of studies utilizing this technique, employ a combination of metaheuristic algorithms such as genetic algorithms and neural networks, which presents several challenges and disadvantages, including the fact that these methods are incredibly time-consuming and frequently produce a locally optimal solution for large data sets with a large number of features. Additionally, when selecting the appropriate feature set, the probability of selecting a subset with redundancy increases due to a lack of consideration for the correlation between the input features. Furthermore, these methods do not prioritize features.

As a result, two distinct dimensional techniques were used and evaluated concurrently in this study to address these issues. While the correlation between each feature and the model output is considered, the correlation between the input features is also examined. Also, features are prioritized, and, to minimize cost and computational load, the bare minimum features necessary for the stock index forecasting model have been selected.

2. MATERIALS AND METHODS

In this study a method called Mutual Information Difference (MID) is used to estimate the Minimum Redundancy Maximum Relevance (mRMR) for selecting features. While considering the correlation of each feature with the model's output, this statistical method also examines the correlation of the input features with one another, and then the features are prioritized. Extracting features is accomplished using principal component analysis (PCA). The prediction model is then based on Support Vector Regression (SVR). It is worth noting that the k-fold cross-validation method (k = 10) was used to divide the data into two categories for training and testing. This method allows the simultaneous use of all data for training and testing, which improves the model's accuracy and usefulness in practice. Moreover, the MAE, MSE, and RMSE criteria were used to assess the performance of these models.

3. RESULTS AND DISCUSSION

The analysis of dimensional reduction techniques confirmed that the feature selection method outperformed feature extraction in selecting the input variables for the SVR model. In this way, it was concluded to focus on the technique of selecting suitable features for the pre-processing of the input variables of the stock index prediction model. Finally, based on the MID results for prioritizing the features affecting stock index forecasting, an algorithm called Initial Selected Features_ Mutual Information Difference was proposed for selecting the appropriate stock index forecast model features. The results of this study's three methods are depicted in Figure 1.

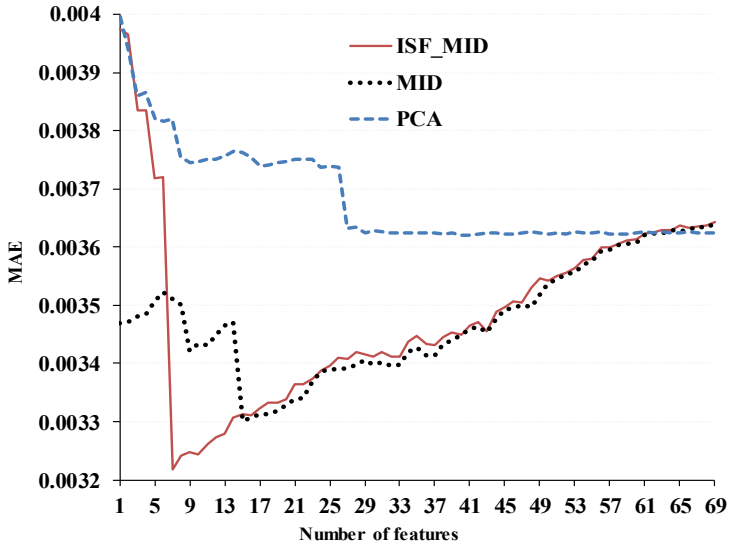

Figure 1. Comparison of PCA, MID, and ISF-MID to select appropriate features for the SVR model in order to forecast the Stock Index

4. CONCLUSION

Comparing the available diagrams demonstrates that the SVR model with features selected using the proposed ISF MID algorithm, which employs only 7 features, produced the best result and accuracy. At the same time, the SVR model with features selected by the MID algorithm (15 features) and features extracted by the PCA algorithm (29 features) had the highest accuracy. In Table 1, the results of the three algorithms' use of the selected features are summarized. According to the results in this table, the methodology used in this study to identify, prioritize, and select appropriate features for the forecasting model in terms of its simplicity and effectiveness of use can be beneficial for researchers in the modeling field because it can be applied to a variety of modeling areas, including the capital market, foreign exchange market, and etc. Additionally, this method can be used for text mining or selecting input features for other smart techniques.

Table 1. Results of using ISF_MID, MID, and PCA in selecting SVR model inputs

PCA	MID	ISF_MID	
29	15	7	Selected features
0.0036	0.0033	0.0032	MAE
0.000036	0.000029	0.000028	MSE
0.0059	0.0053	0.0052	RMSE

Keywords: Dimension Reduction Technique, Feature Selection, Principal Component Analysis, Stock Index Prediction, Support Vector Regression.
JEL Classification: C38, C53, C61, G10, G17.

References

Cavalcante, R. C., Brasileiro, R. C., Souza V. L.F., Nobrega, J. P. & Oliveira A. L. I. (2016). "Computational Intelligence and Financial Markets: A Survey and Future Directions," *Expert Systems with Applications*. 55.194-211.

Zhong, X. & Enke, D. (2017). "Forecasting daily stock market return using dimensionality reduction." *Expert Systems with Applications*, 67, 126-139.