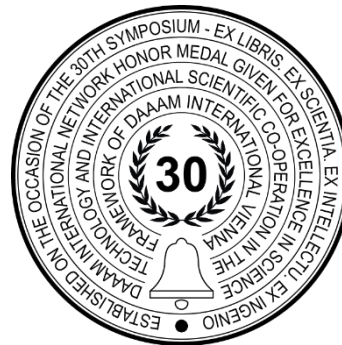# THE EFFECT OF FEATURE SELECTION ON THE PERFORMANCE OF LONG SHORT-TERM MEMORY NEURAL NETWORK IN STOCK MARKET PREDICTIONS

Ive Botunac, Ante Panjkota & Maja Matetic

Abstract

Stock market predictions are a difficult and challenging task affected by numerous interrelated economic, political and social factors caused by non-linear and often unstable movements. Precisely due to the stated nature of financial time series, there is a need to develop advanced systems for stock market prediction. This research seeks to solve one of the problems of such systems, which is reflected in the selection of features to improve the performance of models that are an integral part of the system. In the paper, the wrapper method - recursive feature elimination and the filter method - feature importance, are used for feature selection. A forecasting model based on the long short-term memory (LSTM) neural network was defined to predict the movement of the stock's closing price. With this research we can conclude that for each selected stock there are certain features that have an impact on the results and that it is therefore necessary to carry out the selection of features individually.

Keywords: stock market; machine learning; feature selection; neural network; LSTM

1. Introduction

In the age of economic globalization and the development of computer technologies, the availability of financial data is increasing, which in turn increases the interest in trading in the stock market. Such a rapidly growing availability and amount of data far exceeds a human ability to manually analyse them thus opening up the need to find alternative solutions that could provide an answer to this task. Financial time series data are more complex than other statistic data due to long-term trends, cyclical variations, seasonal variations, and nonlinear movements. They are significantly influenced by many external factors, such as many interrelated economic, political, social, and even the behaviour of the investor himself [1]. The continuous growth of such fluctuating and irregular data has created the need to develop automated systems [13], [14] for efficient analysis in order to be able to extract certain statistical indicators and samples from them. Predicting the future price or the direction of a stock is crucial for investors because it can reduce the risk when making a trading decision. Such decision-making approaches are based on machine learning methods to detect appropriate patterns from available financial time series data and thus generate predictions of the future price or trend of the selected stock.

This research seeks to solve one of the problems of such systems, which is reflected in the feature selection to improve the performance of models that are an integral part of the system. In order to be able to successfully conduct research, it is important to define a methodological approach that basically consists main two parts. The first part presents the features selection while the second part the model for predicting the target variable based on the selected features. The aim of this research is to prove that the features selection can improve the results of the prediction model so that trading decisions can be made on the basis of this prediction model with as low risk per investor as possible.

2. Related work

When conducting research in the field of capital market prediction, we need to solve the problem of selecting input features that will be used in predicting future values [15]. As an example, according to the authors in [2] and [3] we find the use of methods such as the method of recursive feature elimination (RFE) to solve this problem. Numerous studies use different machine learning methods to predict the return of investment in the stock market or to predict the direction of movement. One common thing among the researches is the use of some of the technical indicators that are the basis in conducting technical analysis in the stock market.

In the research presented in [2] we can see that the author uses the method of recursive feature elimination (RFE) to select features. The use of technical indicators in combination with additional online data sources (Google search data) in the research showed greater predictive power than any of these sources alone. Using decision trees, neural networks and a support vector machine, has up to 85% accuracy in predicting the direction of movement the next day for, in this case, the AAPL stock (Apple).

In addition to implementing the feature selection methods to reduce dimensionality and improve the model performance in [5], we can see that by applying empirical wavelet transform, the author achieves a better decompression effect on complex stock market price series. Also, in the previously conducted research [6] and with discrete wavelet transformation as a technique for data preprocessing, better results were obtained in predicting the future trend of stock movements.

3. Methodology

In order to provide a solution to the set problem, we use an approach that consists of two parts, where in the first part we handle feature selection and then in the second part we make predictions based on these features. An important role in achieving the goal of this research is played by the correct selection of input features using techniques based on machine learning methods [2]. By properly implementing feature selection, we ensure that the performance of the prediction model is improved.

3.1. Data set description

When developing a prediction model, one of the most important factors is the raw data from which we generate input features which are then divided into a training and test set. To conduct this research, data from the financial time series in the period from January 1, 2015 to December 31, 2019 were collected from Alpha Vantage API from Apple (AAPL), Microsoft (MSFT) and Facebook (FB) stocks. In Figure 1 we can see the close price for the Apple stock in the aforementioned period.
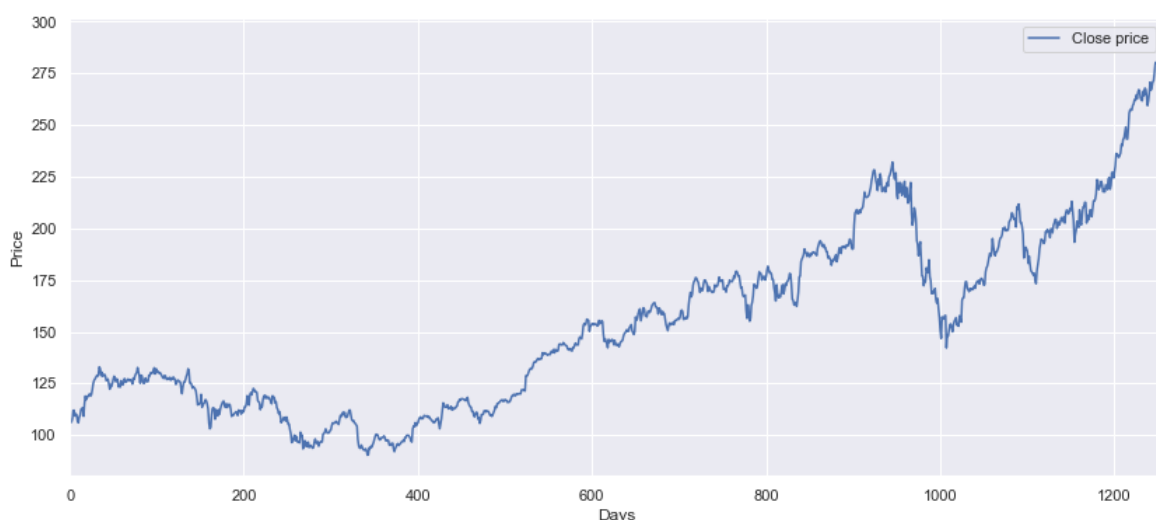

Fig. 1. Display of AAPL stock closing price

The data collected include a total of 14 different features made up of standard financial indicators of the financial time series and of technical indicators [4] shown in Table 1.

No.	Feature name	Label
0.	Closing Price	C
1.	Opening Price	O
2.	High Price	H
3.	Low Price	L
4.	Volume	V
5.	Simple 10-Day Moving Average	SMA
6.	Weighted 10-Day Moving Average	WMA
7.	Momentum	MOM
8.	Moving Average Convergence Divergence	MACD
9.	Stochastic K%	K%
10.	Stochastic D%	D%
11.	Relative Strength Index	RSI
12.	Williams %R	%R
13	On-Balance Volume	OBV

Table 1. Features overview

3.2. Feature selection

Feature selection is the process by which we reduce the number of input features in our prediction model. By reducing the number of features, we ensure faster execution of the prediction model while improving its performance [7] since the model uses only those features that have been shown to have the greatest impact on the dependent variable, which in this case is the future closing price.

As we solve the problem with the application of supervised learning in this research, we use the wrapper and filter methods to select features. For the wrapper method we use recursive feature elimination (RFE) with linear regression (LR) while for the filter method we use feature importance (FI) with linear regression (LR), decision tree regression (DTR) and random forest regression (RFR). The best feature subset is selected for a particular closing price stock prediction by comparing the mentioned feature selection methods.

3.3. Prediction model

For the prediction model in this research, we have selected LSTM neural network based on numerous studies [8], [9], [10] in this area where it has been shown that this machine learning method achieves significantly better results than other machine learning methods. During the experiment, the hyperparameters of the LSTM neural network were adjusted in order to improve the quality of the model resulting with a better prediction.

4. Experimental procedure, results and discussion

To begin conducting the experiment of this research it is important to preprocess data as the first step to be able to transform the data into an applicable form which can be used to conduct feature selection and prediction. In the development of the prediction model, collected data were divided into training and test set.

4.1. Feature scaling

We perform data normalization to scale their values within the given ranges, which in this case include values from -1 to 1. We use described data processing technique in order to avoid large values or large deviations between different features.

4.2. Feature selection

After the feature scaling, we proceed to the feature selection process. We come to unexpected outcomes when implementing the recursive feature elimination (RFE) method using linear regression over selected stocks. Figure 2 shows the box plot graph of the results in measuring negative mean absolute error (NMAE) related to the number of features over the AAPL stock.

From the graphical presentation in the form of a box plot, the Occam principle can suggest two features as the simplest solution. Nevertheless, that is less likely, and further investigation needs to be carried out, which is out of scope for this preliminary study.
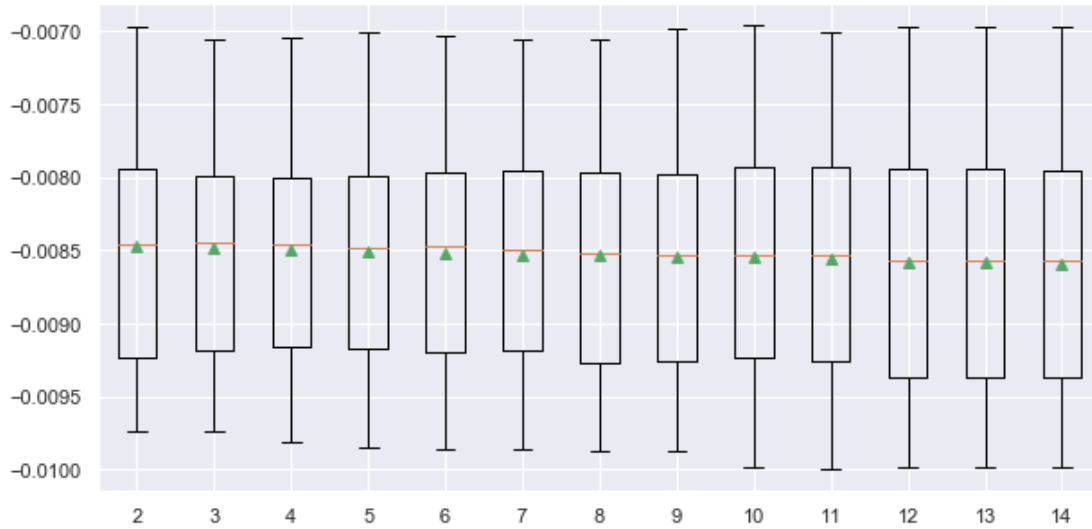

Fig. 2. Box plot of the selected number of features and NMAE using the RFE method with linear regression on AAPL stock

We obtained ambiguous initial results from the RFE method in feature selection, which directs this preliminary study toward the feature importance method. This method shows more promising results in solving the problem of features selecting for predicting future stock movements.

With these results, we can notice that there is a certain pattern in the features that are selected and the features that are not selected. Thus, for example, with all the methods used in feature importance, we see that the closing price (0. C) proved to be a selected feature while the volume of trading (4. V) never proved to be a selected feature. Also, some of the technical indicators like simple moving average (5. SMA) and weighted moving average (6. WMA) proved to be more often selected features than technical indicators stochastic K% (9. K%), stochastic D% (10. % D) and the relative strength index (11. % R).

Company	Feature													
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
Apple (AAPL)	**0.8636**	**0.0872**	-0.0826	**0.0433**	0.0038	**0.0805**	-0.0001	**0.0099**	**0.0105**	-0.0004	-0.0019	-0.0077	**0.0084**	0.0002
Microsoft (MSFT)	**0.5861**	**0.0246**	-0.0339	**0.2043**	0.0008	**0.0222**	0.1936	-0.0017	-0.0080	-0.0031	**0.0047**	**0.0078**	0.0012	0.0035
Facebook (FB)	**0.7861**	**0.0208**	**0.1697**	**0.0626**	-0.0258	**0.0938**	-0.1502	-0.0194	**0.0138**	0.0051	-0.0098	-0.0155	**0.0154**	**0.0097**

Table 2. Overview of the results for the feature importance method using the Linear Regression method

Company	Feature													
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
Apple (AAPL)	**0.7579**	0.0004	0.0041	**0.1221**	0.0002	**0.0276**	**0.0856**	0.0001	0.0002	0.0003	0.0001	0.0006	0.0002	0.0002
Microsoft (MSFT)	**0.0181**	0.0030	**0.1120**	**0.0685**	0.0001	**0.0093**	**0.7879**	0.0001	0.0005	0.0001	0.0001	0.0001	0.0001	0.0001
Facebook (FB)	**0.9807**	0.0002	0.0005	**0.0152**	0.0003	0.0001	0.0004	0.0002	0.0004	0.0002	0.0001	0.0006	0.0001	0.0004

Table 3. Overview of the results for the feature importance method using the Decision Tree Regression method

Company	Feature													
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
Apple (AAPL)	**0.75 79**	0.00 04	0.00 41	**0.12 21**	0.00 02	**0.02 76**	**0.08 56**	0.00 01	0.00 02	0.00 03	0.00 01	0.00 06	0.00 02	0.00 02
Microsoft (MSFT)	**0.01 81**	0.00 30	**0.11 20**	**0.06 85**	0.00 01	**0.00 93**	**0.78 79**	0.00 01	0.00 03	0.00 01	0.00 01	0.00 01	0.00 01	0.00 01
Facebook (FB)	**0.98 07**	0.00 02	0.00 05	**0.01 52**	0.00 03	0.00 01	0.00 04	0.00 02	0.00 04	0.00 02	0.00 01	0.00 06	0.00 01	0.00 04

Table 4. Overview of the results for the feature importance method using the Random Forest Regression method

4.3. Hyperparametar tuning

In order to improve the performance of the prediction model, we perform selection of the best parameters that represent the number of neurons in each layer of the neural network and the number of training epochs, adjusting the value of drop optimization technique and value of Adaptive Moment Estimation (ADAM) optimization algorithm. In the adjustment process itself, we use an approach called batch normalization to speed up the training, and the dropout technique to prevent the possibility of overfitting [10].

We also use Adaptive Moment Estimation (ADAM) as the chosen optimization algorithm. Using the grid search technique in the process of training the prediction model, we use different values of these parameters starting from the lowest with a gradual increase [11]. With this technique of searching, i.e. optimizing parameters, we train the model through all possible combinations in the predefined subset of values shown in table 5. Table 5. Also shows the selected parameter values.

Hyperparametar	Predefined subset of values	Selected value
First layer (LSTM cell)	64, 128, 256, 512	512
First dropout	0.1, 0.2, 0.3, 0.4	0.1
Second layer (LSTM cell)	64, 128, 256, 512	512
Second Dropout	0.1, 0.2, 0.3, 0.4	0.1
Third layer (dense cells (ReLu))	8, 16, 32, 64	64
Four layer (dense cells (ReLu))	1	1
Adam	0.1, 0.2, 0.3, 0.4	0.1
Batch	64, 128, 256, 512	128
Epoch	100, 200, 300, 400, 500	300

Table 5. Overview of the prediction model architecture and parameters

4.4. Prediction model performance result

Firstly, research results show which combinations of selected features achieve the best results. As this study is a regression prediction, mean absolute error (MAE) and mean square error (MSE) are used for model performance achievements measures [12]. In Table 6, we can see a comparison of the results with differently selected input features for three different stocks according to the method used when conducting the features selection.

From the results of Table 6., according to performance achievements measures on the test set, we can see that each stock achieves better results with a different set of selected features and different methods used when selecting features. We can see that the feature selections greatly affect the LSTM prediction results, and it's specific to all selected stocks (see tables 2, 3, and 4).

Figure 3 compares the actual and predicted closing price at the MSFT stock test set using the feature importance method from decision tree regression (DTR).

Company	Feature importance selection + LSTM					
	Linear Regression (LR) + LSTM		Decision Tree Regression (DTR) + LSTM		Random Forest Regression (RFR) + LSTM	
	MAE	MSE	MAE	MSE	MAE	MSE
Apple (AAPL)	0.01637	0.00046	0.01598	0.00041	**0.01547**	**0.00041**
Microsoft (MSFT)	0.02101	0.00061	**0.01800**	**0.00046**	0.02649	0.00092
Facebook (FB)	**0.01606**	**0.00046**	0.01655	0.00049	0.01669	0.00049

Table 6. Overview of the result of the LSTM prediction model

Fig. 3. Presentation of the true closing price and the projected closing price of the MSFT stock

5. Conclusion

In this study, several different methods for selecting features were tested to predict stock market movements. When conducting feature selection, the recursive feature elimination and feature importance methods were used, while the LSTM neural network was used to make predictions.

This preliminary study showed that the selection of features has an impact on the performance of the prediction model. For the three selected stocks over which the feature selection process was performed, we achieve different results by the used feature importance model. After all, we can conclude that different features could exist for each selected stock that significantly impacts the LSTM prediction results. Therefore, it is a necessity for further research over the effect of the selection of features with particular stocks.

This research opens up space for numerous improvements to the used forecasting model. One of the ideas for future research is the development of automated trading systems on electronic exchanges using machine learning techniques like reinforcement or Q learning. With the current results obtained from this forecasting model, a number of trading strategies can be implemented in the form of generating buy and sell signals in order to facilitate the decision-making process for traders. Machine learning with application in the field of trading in the stock market opens up an entirely new domain that has just begun its development.

6. Acknowledgments

7. References

[1] Dash, R. & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques, The Journal of Finance and Data Science, Vol. 2, No. 1, pp. 42–57, https://doi.org/10.1016/j.jfds.2016.03.002

[2] Weng, B.; Ahmed, M. A. & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. Expert Systems with Applications, Vol. 79, pp. 153–163. https://doi.org/10.1016/j.eswa.2017.02.041

[3] Weng, B.; Lu, L.; Wang, X.; Megahed, F. M. & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. Expert Systems with Applications, Vol. 112, pp. 258–273. https://doi.org/10.1016/j.eswa.2018.06.016

[4] Li, X.; Wu, P. & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing and Management, Vol. 57, No. 5, https://doi.org/10.1016/j.ipm.2020.102212

[5] Liu, H. & Long, Z. (2020). An improved deep learning model for predicting stock market price time series. Digital Signal Processing: A Review Journal, Vol. 102, 102741. https://doi.org/10.1016/j.dsp.2020.102741

[6] Botunac, I.; Panjkota, A. & Matetic, M. (2019). The importance of time series data filtering for predicting the direction of stock market movement using neural networks. Proceedings of the 30th DAAAM International Symposium, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-22-8, ISSN 1726-9679, Vienna, Austria, https://doi.org/10.2507/30th.daaam.proceedings.xxx

[7] Weng, B. (2017). Application of machine learning techniques for stock market prediction, Ph.D. Dissertation, Graduate Faculty of Auburn University.

[8] Fischer, T. & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, Vol. 270, No. 2, pp. 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

[9] Long, W.; Lu, Z. & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems, Vol. 164, pp. 163–173. https://doi.org/10.1016/j.knosys.2018.10.034

[10] Oncharoen, P. & Vateekul, P. (2018). Deep Learning for Stock Market Prediction Using Event Embedding and Technical Indicators. ICAICTA 2018 - 5th International Conference on Advanced Informatics: Concepts Theory and Applications, pp. 19–24. https://doi.org/10.1109/ICAICTA.2018.8541310

[11] Chalvatzis, C. & Hristu-Varsakelis, D. (2019). High-performance stock index trading: making effective use of a deep LSTM neural network. pp. 1–24. http://arxiv.org/abs/1902.03125

[12] Henrique, B. M.; Sobreiro, V. A. & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. Expert Systems with Applications, Vol. 124, pp. 226–251. https://doi.org/10.1016/j.eswa.2019.01.012

[13] Park, H.; Sim, M. K., & Choi, D. G. (2020). An intelligent financial portfolio trading strategy using deep Q-learning. Expert Systems with Applications, Vol. 158, 113573. https://doi.org/10.1016/j.eswa.2020.113573

[14] Wu, X.; Chen, H., Wang, J., Troiano, L., Loia, V., & Fujita, H. (2020). Adaptive stock trading strategies with deep reinforcement learning methods. Information Sciences, Vol. 538, pp. 142–158. https://doi.org/10.1016/j.ins.2020.05.066

[15] Raschka, S. & Mirjalili, V. (2019). Python Machine Learning, Packt Publishing, ISBN: 978-1-78995-575-0, UK