Croatian Review of Economic, Business and Social Statistics (CREBSS)
UDK: 33;519,2; DOI: 10.1515/crebss; ISSN 1849-8531 (Print); ISSN 2459-5616 (Online)
Vol. 6, No. 2, 2020, pp. 4-11

sciendo

Comparing classification algorithms for prediction on CROBEX data

Silvija Vlah Jerić
Faculty of Economics & Business of University of Zagreb, Croatia,
svlah@net.efzg.hr

Abstract
The main objective of this analysis is to evaluate and compare the various classification algorithms for the automatic identification of favourable days for intraday trading using the Croatian stock index CROBEX data. Intra-day trading refers to the acquisition and sale of financial instruments on the same trading day. If the increase between the opening price and the closing price of the same day is substantial enough to earn a profit by purchasing at the opening price and selling at the closing price, the day is considered to be favourable for intra-day trading. The goal is to discover relation between selected financial indicators on a given day and the market situation on the following day i.e. to determine whether a day is favourable for day trading or not. The problem is modelled as a binary classification problem. The idea is to test different algorithms and to give greater attention to those that are more rarely used than traditional statistical methods. Thus, the following algorithms are used: neural network, support vector machine, random forest, as well as k-nearest neighbours and naïve Bayes classifier as classifiers that are more common. The work is an extension of authors' previous work in which the algorithms are compared on resamples resulting from tuning the algorithms, while here, each derived model is used to make predictions on new data. The results should add to the increasing corpus of stock market prediction research efforts and try to fill some gaps in this field of research for the Croatian market, in particular by using machine learning algorithms.

Keywords: classification algorithms, CROBEX, day trading, stock market prediction.

JEL classification: C38, C45, G170.
DOI: 10.2478/crebss-2020-0007

Received: October 31, 2020
Accepted: November 17, 2020

Introduction
There is a lot of research in the literature on predicting stock market price movements. Many of them use machine learning methods as an alternative to classical statistical methods. Of these methods, for example, neural networks, decision trees, and support vector machines are often used successfully. However, it seems that similar studies are rare in Croatia, as noticed by, for example, Šego and Škrinjarić (2018). This paper should therefore contribute in this regard.

Some of the many research topics in the field of stock market price forecasting are regression problems, in which one tries to predict the stock price or the value of the stock index, and classification problems, in which one tries to predict the direction of the changes of these values, for example. Some of the first research in this area used neural networks to predict the Tokyo stock market (Mizuno, Kosaka, Yajima, Komoda, 1998; Kimoto, Asakawa, Yoda, Takeoka, 1990). Some other similar studies have used the Bayes classifier (Pop, 2006; Shin, Kil, 1998; Tsaih, Hsu, Lai, 1998;) and support vector machines (Ince, Trafalis, 2007; Moreira, Jorge, Soares, Sousa, 2006). Also, many studies compare the performance of different methods such as neural networks, support vector machines, k-nearest neighbours, naïve Bayes classifier, genetic algorithms, decision trees etc. (Kara, Boyacioglu, 2011; Qi, Zhang, 2008; Qian, Rasheed, 2007; Zemke, 1999). There are also some papers on using machine learning algorithms for predicting stock market trends on Zagreb Stock Exchange (ZSE) such as research by Manojlović and Štajduhar (2015) who use random forest.

The aim of this paper is to compare five algorithms according to their success in classifying days as favourable or unfavourable for trading within one day e.g. intra-day trading on the Croatian stock market. Namely, intra-day trading consists of buying and selling financial instruments within one trading day and a day is considered favourable for trading if the difference between the opening price and the closing price on the same day is large enough, while otherwise it is unfavourable. In this research, the price of the Croatian stock index CROBEX is observed, and the selected algorithm should automatically recognize whether the following day will be favourable or unfavourable for intra-day trading depending on the value of selected technical indicators for the current day. The algorithms used are neural networks (NN), support vector machines (SVM), random forest (RF), k-nearest neighbours (KNN) and naïve Bayes classifier (NB).

This paper is a continuation of the research from Vlah Jerić (2020a, b) in which a preliminary comparison of algorithms based on resamples resulting from tuning the algorithms was performed and the possibility of extracting decision rules from the obtained RF model was examined. However, unlike previous work in which the algorithms are compared only on resamples, in this work the analysis goes deeper and each derived model is used to make predictions on new data. This should give better insights into how the algorithms really perform in terms of accuracy.

The next, second, chapter describes the data and algorithms that are analysed in the research, while the third chapter presents the results of the research. The last, fourth chapter, provides conclusions and guidelines for future research.

The data and selected algorithms

The data description

The data used consists from daily opening, closing, high and low values of Croatian stock index CROBEX for the last ten years, more precisely from the beginning of year 2010 until the end of year 2019. As the first and most important Zagreb Stock Exchange equity index, CROBEX is perceived as an indicator of the movement of Croatian stock market prices. There are 2492 records (observations) to start with since each data record refers to one trading day.

In this research, the day is considered to be favourable if the increase between the opening price and the closing price of the same day is larger than 0.3%, while otherwise, the day is considered to be unfavourable for intra-day trading. This benchmark value could be set at any chosen value and here it is 0.3 as in Bruni (2017) who suggested it should provide a reasonable opportunity for profit. In such way, the class is assigned to each daily record and the classification algorithms' task will be to try to guess whether the following day will be favourable by looking at values of a set of technical indicators.

The indicators are chosen following Bruni (2017) where detailed explanations of the calculation procedure can be found. Values of the parameters used in the calculations were chosen as values widely used by traders and scientists as defaults in intraday trading. A total of 18 indicators are used:

- Momentum over five periods (MOM),
- an exponential moving average (EMA) over 12 and 26 periods (EMA12 and EMA26, respectively),
- Moving Average Convergence/Divergence with 12, 26 and 9 time periods (respectively) selected as three parameters needed for calculation (MACD),
- Return On Investment over 10, 20 and 30 periods (ROI10, ROI20 and ROI30 respectively),
- Relative Strength Index over 10, 14 and 30 periods (RSI10, RSI14 and RSI30 respectively),
- Stochastic Relative Strength Index over 10, 14 and 30 periods (SRSI10, SRSI14 and SRSI30, respectively),
- Average True Range over 14 periods (ATR),
- Average Directional Index over the last 14 periods (ADX),
- Williams %R over 14 periods (WPR),
- Commodity Channel Index over 20 periods (CCI),
- Ultimate Oscillator with 7, 14 and 28 time periods (respectively) selected as three parameters needed for calculation (UO).

During the calculations of the class and the indicators, 40 data records are lost. The total number of data records used in the experiments is therefore 2452. Among these, there are 657 favourable days and 1795 unfavourable days, which means the data is imbalanced since there are far more unfavourable days than the favourable ones. Table 1 shows summary statistics for the selected indicators.

Table 1 Summary statistics for the selected indicators

Indicator	Min	Max	Mean	St.dev.
MOM	-237.10	179.18	0.20	29.48
EMA12	1594.33	2307.58	1856.90	151.84
EMA26	1601.54	2295.04	1856.44	148.07
MACD	-5.00	3.47	0.01	0.95
ROI10	$-8.02 \cdot 10^{-6}$	$4.82 \cdot 10^{-6}$	$-3.68 \cdot 10^{-10}$	$1.20 \cdot 10^{-6}$
ROI20	$-4.57 \cdot 10^{-6}$	$3.20 \cdot 10^{-6}$	$-8.50 \cdot 10^{-10}$	$9.23 \cdot 10^{-7}$
ROI30	$-4.33 \cdot 10^{-6}$	$2.42 \cdot 10^{-6}$	$-2.66 \cdot 10^{-9}$	$7.97 \cdot 10^{-7}$
RSI10	7.77	84.03	49.58	15.96
RSI14	11.34	79.68	49.57	13.91
RSI30	17.66	72.56	49.58	10.24
STOCHRSI10	0.00	1.00	0.49	0.38
STOCHRSI14	0.00	1.00	0.49	0.38
STOCHRSI30	0.00	1.00	0.49	0.39
ATR	8.41	73.39	19.15	8.46
ADX	6.54	78.11	26.54	13.21
WPR	0.00	1.00	0.49	0.30
CCI	-294.16	392.65	-0.28	115.07
UO	30.77	70.00	52.00	5.42

Source: authors' calculations.

The selected algorithms and testing scheme

Five types of algorithms that can be used for classification tasks were selected to try to classify the days as favourable or unfavourable: neural network (Ripley, 2007; Haykin, 1999; Bishop, 1995); support vector machines (James, 2013; Kuhn and Johnson, 2013; Hastie, Tibshirani and Friedman, 2001); random forest (Izenman, 2008; Hastie, Tibshirani and Friedman, 2001; Breiman, 2001; Ho, 1998); k-nearest neighbours (Hall, Park and Samworth, 2008; Hastie, Tibshirani and Friedman, 2001); naïve Bayes classifier (Murty and Devi, 2011; Russell, 2010).

The algorithms were evaluated by their performance in terms of their successfulness in classifying days as favourable or unfavourable correctly. However, instead of the most common metric used for this – accuracy, Cohen's kappa (Cohen, 1960) is chosen as performance assessment metric because it is better suited for imbalanced data as it reveals how much better a chosen classifier performs than a classifier that merely guesses randomly according to the frequency of each class. Namely, since accuracy is calculated as the proportion of true results among the total number of cases examined, it is not a representative measure for evaluation of classification algorithms performance on data which are imbalanced or skewed. For imbalanced classification, a high accuracy can be attained by only predicting the majority class, i.e. in the case of this CROBEX data, predicting that all days were unfavourable would give high accuracy. However, there is also a downside to using Cohen's kappa: unlike simple accuracy, it is more difficult to interpret and understand.

For the purposes of proper evaluation of the selected algorithms performances, the data was dived into ten sections (equal length arrays of consecutive records) i.e. time slices. Each of the time slices was further divided on training set and test set such that the training set part consisted of 80% of the consecutive records in that time slice, while the test set section was the rest 20%. Training set was used to obtain an optimal model for the given algorithmic approach and then this model was used on new data (data that was not used in selecting the optimal model) i.e. the test set. By doing so, the models are cross-validated and their hyper parameters are tuned for optimal model performance and then tested on unused data. Within this cross-validation scheme, the optimal model mentioned above was selected following the same cross validation technique, but only on the training data for the given time slice. That means that each training set was further divided into ten parts i.e. ten time slices where the cross validation is used to select the optimal model.

The described type of data splitting for cross validation that basically moves the training and test sets in time used in this research (for tuning the algorithms and selecting the optimal model, as well as for evaluating its' performance on new data) is known as rolling forecasting origin technique (Hyndman & Athanasopoulos, 2020). There are different forms of this technique and the one that is used in this analysis is the one where the training set moves in time meaning that it always contains the same number of records instead of always beginning from the first sample with the set size varying over data splits. Also, the predictions are made for more than only one period ahead, making it harder for predicting accurately.

Results

For class prediction on CROBEX data, the simplest type of artificial neural network devised was used – the feed-forward neural network. More precisely, the feed-forward NN with a single hidden layer is fitted, after being tuned by varying the number of hidden units and weight decay. The activation function is logistic. As for the SVM, the CROBEX data were processed by support vector machine with radial basis function

(RBF) kernel which is tuned over the cost parameter and the RBF kernel parameter sigma. On the other hand, RF is tuned by varying the number of randomly selected predictors i.e. the number of variables randomly sampled as candidates at each split. For all three previously mentioned algorithms, the amount of granularity in the tuning parameter grid is set to ten. Also, in this research, the best choice of k in KNN, which depends upon the data, is tuned by trying five different values. Finally, on CROBEX data, NB classifier is tuned distribution type – parameter that allows adjusting the bandwidth of the kernel density and bandwidth adjustment (zero to five).

R (R Core Team, 2020) was used for running all of the experiments on CROBEX data, as well as the statistical tests that followed. The summary of the results on running all of the selected algorithms in order to evaluate their performance on unknown data based on the cross-validation scheme described in the previous chapter are given in table 2.

Table 2 Summary of Kappa values on new data

	NN	SVM	RF	KNN	NB
Min.	-0.082760	-0.056289	-0.023428	-0.28699	-0.113845
1st Qu.	-0.073242	-0.001409	-0.005921	-0.02764	-0.008397
Median	-0.039208	0.000000	0.011913	-0.01020	0.016697
Mean	-0.039049	0.011771	0.016468	-0.03252	0.018084
3rd Qu.	-0.004842	0.017387	0.037218	0.01024	0.035932
Max.	0.002756	0.133972	0.061990	0.06160	0.151254

Source: authors' calculations.

It should be noted that the Table 2. shows the Kappa values of applying the optimal model (obtained through tuning the algorithms) on new data and the auxiliary results on resamples that are used to make the selection of the optimal model of each approach are not included here. Visualisation of these results in form of box-whiskers plots is given in Figure 1.

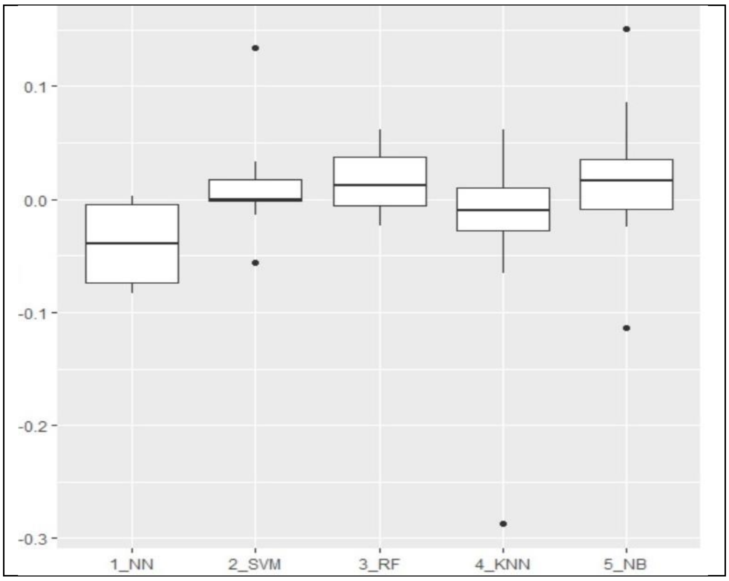

Figure 1 Box-whiskers plots for summary of Kappa values on new data
Source: authors' calculations)

The significance of differences between mean Kappa values for the five algorithms is tested using Friedman test as the non-parametric version of the well-known ANOVA. Namely, when evaluating the performance of machine learning algorithms, the

Croatian Review of Economic, Business and Social Statistics (CREBSS)
UDK: 33;519,2; DOI: 10.1515/crebss; ISSN 1849-8531 (Print); ISSN 2459-5616 (Online)
Vol. 6, No. 2, 2020, pp. 4-11

assumptions which ANOVA is based on are most likely to be violated, so we follow the suggestions of Demšar (2006) for these tests. The Friedman test reports a significant difference between the five evaluated algorithms ($p = 0.0107$) for $\alpha = 0.05$. For pairwise comparisons of the multiple classifiers we use the corresponding Nemenyi post-hoc test (Nemenyi-Wilcoxon-Wilcox all-pairs test) and the results are given in Table 3.

Table 3 Nemenyi post-hoc test for Kappa values

	NN	SVM	RF	KNN
SVM	0.081	-	-	-
RF	0.010	0.955	-	-
KNN	0.790	0.618	0.211	-
NB	0.157	0.999	0.860	0.790

Source: authors' calculations.

From the results in Table 2, it can be seen that NB had the highest (the best) mean and median Kappa values, though the differences between those and of the second-best RF were very small. Indeed, the results in Table 3 show that the difference between the mean Kappa values by NB and RF is not significant which was expected when looking at the box-whiskers plot from Figure 1. On the other hand, a much greater variability for NB than for RF can be observed. Thus, it seems that choosing RF algorithm for this type of classification would make sense and exploring the possibility of rule extraction from RF as a way of adding interpretability is appealing. It should be noted that Table 2 contains a lot of negative values and negative Kappa values mean that classification was pretty bad.

Other than the results used in previously driven conclusions, Table 3 also reports a significant difference between RF and NN algorithm ($p = 0.010$) for $\alpha = 0.05$, as well as a significant difference between SVM and NN algorithm ($p = 0.081$) for $\alpha = 0.1$. It is also interesting to notice that there is no significant difference between KNN and any other method, as well as between NB and any other method. Looking at box-whiskers plots gives an idea for that. Namely, both methods show great variability in their performance. Thus, they do not seem reliable. On the other hand, NN presented itself as a huge disappointment as it was really bad compared to other algorithms. However, it could surely be tuned better to give its best and show that it can perform much better. Still, the idea here is to compare algorithms so tuning them too much would not make much sense. This way, they were tuned to a certain extent but without overthinking it and thus by setting the same tuning grid granularity where possible.

In the end, although accuracy metric has bad representativeness of imbalanced classification performance, it is interesting to mention, due to its popularity, that the mean accuracy values for all approaches ranged from 0.6370 to 0.7120, which is reasonable.

Conclusion

This paper compared the performance of five different classification algorithms for the automatic detection of favourable days for intraday trading using the data from Croatian stock index CROBEX. Greater attention was given to the methods that are more rarely used than traditional statistical methods. The findings contribute to this area of research for the Croatian market where there are not many such analyses, in particular by using machine learning algorithms, and also to the growing corpus of stock market prediction analysis efforts in general.

Evaluations of neural network, support vector machine, random forest, k-nearest neighbours and naïve Bayes classifier showed significant difference in performance on new data according to Cohen's kappa. Random forest seems to be promising for this type of classification and thus it would make sense to explore the possibility of rule extraction from RF as a way of adding interpretability, which is one of the possible research directions for future research. Although some preliminary research in this topic has been conducted, it needs much further analysis still as the question of how much of accuracy would be sacrificed that way remained open.

Also, it would be interesting to see how these algorithms behave on other stock markets in the region as well as to investigate is there any pattern in the parameters chosen in the optimal models. Finally, although CROBEX, being a stock index, cannot be bought and sold, researching the algorithms for predictions of CROBEX data could be applied for trading on Croatian stock market by using these procedures on actual stocks.

References

1. Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.
2. Breiman, L. (2001). Random Forests. *Machine Learning*, Vol. 45, No. 1, pp. 5-32.
3. Bruni, R. (2017). Stock Market Index Data and indicators for Day Trading as a Binary Classification problem. *Data in Brief*, Vol. 10, pp. 569-575.
4. Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *The Journal of Machine Learning Research*, Vol. 7, pp. 1-30.
5. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37-46.
6. Hall, P., Park, B. U., Samworth, R. J. (2008). Choice of neighbor order in nearest-neighbor classification. *Annals of Statistics*, Vol. 36, No. 5, pp. 2135-2152.
7. Hastie, T., Tibshirani, R., Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
8. Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
9. Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 8, pp. 832-844.
10. Hyndman, R. J., Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, Melbourne.
11. Ince, H., Trafalis, T. B. (2007). Kernel principal component analysis and support vector machines for stock price prediction. *IIE Transactions*, Vol. 39, No. 6, pp. 629-637.
12. Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Springer, New York.
13. James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer, New York.
14. Kara, Y., Boyacioglu, M. A., Baykan, O. K. (2011). Predicting direction of stock Price index movement using artificial neural networks and support vector machines: the sample of the Istanbul stock exchange. *Expert Systems with Applications*, Vol. 38, pp. 5311-5319.
15. Kimoto, K., Asakawa, K., Yoda, M., Takeoka, M. (1990). Stock market prediction system with modular neural network. In *Proceedings of the International Joint Conference on Neural Networks*, San Diego.
16. Kuhn, M., Johnson, K. (2013). *Applied predictive modeling*. Springer, New York.
17. Manojlović, T., Štajduhar, I. (2015). Predicting Stock Market Trends Using Random Forests: A Sample of the Zagreb Stock Exchange. In *Proceedings of MIPRO CIS - Intelligent Systems Conference*.
18. Mizuno, H., Kosaka, M., Yajima, H., Komoda, N. (1998). Application of Neural Network to Technical Analysis of Stock Market Prediction. *Studies in Informatic and Control*, Vol. 7, No. 3, pp. 111-120.

19. Moreira, J. M., Jorge, A. M., Soares, C., Sousa, D. J. F. (2006). Improving SVM-Linear predictions using CART for example selection. In *Proceedings of the 16th International Conference on Foundations of Intelligent Systems*, Springer-Verlag.
20. Murty, M. N., Devi, V. S. (2011). *Pattern Recognition: An Algorithmic Approach*. Springer-Verlag, London.
21. Shin, F. B., Kil, D. H. (1998). Classification Cramer–Rao bounds on stock price prediction. *Journal of Forecasting*, Vol. 17, No. 5-6, pp. 389-399.
22. Qi, M., Zhang, G. P. (2008). Trend time series modeling and forecasting with neural networks. *IEEE Transactions on Neural Networks*, Vol. 19, No. 5, pp. 808-816.
23. Qian, B., Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, Vol. 26, No. 1, pp. 25-33.
24. R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
25. Ripley, B. D. (2007). *Pattern Recognition and Neural Networks*. Cambridge University Press.
26. Russell, S. J. (2010). *Artificial intelligence: a modern approach*. Prentice Hall, Upper Saddle River.
27. Šego, B., Škrinjarić, T. (2018). Kvantitativna istraživanja Zagrebačke burze - pregled istraživanja od osnutka do 2018. godine. *Ekonomski pregled*, Vol. 69, No. 6, pp. 655-743.
28. The Zagreb Stock Exchange (2020). *Index page*. Available at www.zse.hr [13 January 2020].
29. Vlah Jerić, S. (2020a). Evaluation of Alternative Approaches in Classification Algorithms for Prediction of Stock Market Index: Case of Crobex. In *Recent Applications of Financial Risk Modelling and Portfolio Management*, Škrinjarić, T., Čižmešija, M., Christiansen, B. (Eds.), IGI Global, USA.
30. Vlah Jerić, S. (2020b). Rule Extraction from Random Forest for Intra-Day Trading Using CROBEX Data. *Proceedings of 11th FEB Zagreb International Odyssey Conference on Economics and Business*, pp. 411-419.
31. Zemke, S. (1999). Nonlinear index prediction. *Physica A: Statistical Mechanics and its Applications*, Vol. 269, No. 1, pp. 177-183.

About the author

Silvija Vlah Jerić currently holds a position of an associate professor at Department of mathematics of Faculty of Economics & Business, University of Zagreb. After gaining a BSc degree in Mathematics at Faculty of science, she continued her education on Faculty of Economics & Business Zagreb, where she gained her MSc and PhD degrees. Her field of study is quantitative economics and her main research interests are applications of operational research methods and machine learning in economics. Most of her publications so far involve optimisation methods with special attention to multi-objective optimisation, metaheuristics, machine learning and forecasting evaluations. Author can be contacted at *svlah@net.efzg.hr*.