-Societies and Sustainability-
Open access peer-reviewed journal of SGEM WORLD SCIENCE (SWS) Scholarly Society Vienna

Issue 1/2025

APPLYING OF RANDOM FOREST AND SUPPORT VECTOR MACHINE IN PREDICTING PRICES OF URANIUM COMPANIES

Lukasz Sroka
University of Economics in Katowice, **Poland**

ABSTRACT

Due to the war in Ukraine and restrictions on the export of hydrocarbons from Russia by the European countries, uranium companies are again becoming an interesting sector in terms of investment. Consequently, it is important for investors to have accurate forecasts of uranium sector. This article applies machine learning algorithms such as the Random Forests and the Support Vector Machine to predict future URA ETF prices for the next five periods. The study was conducted using data on the ETF Global X Uranium for the period from 08/11/2010 to 31/05/2023 was obtained from investing.com. The data contains information about the stock such as High, Low, Open, Close, Adjacent close and Volume and several well-known technical indicators. The research showed that both the Random Forest and the Support Vector Machine forecast prices with less bias than the classic ARIMA model. The Random Forest algorithm forecasted prices with a constant level of bias over the forecasting period, while the error of the forecasts calculated by the Support Vector Machine algorithm for the first three periods was the lowest compared to the rest of the analyzed models. Research showed that the Random Forest algorithm and the Support Vector Machine can be used to make correct predictions for uranium sector.

Keywords: uranium; forecasting; machine learning; random forests; support vector machine.

Introduction

Predicting prices for the financial markets is the subject of research by many scientists. There is a lot of discussion in the literature about which model predicts prices with the less bias. This discussion is important because proper prediction of share prices, futures or other financial instruments can significantly help investors in making their financial decisions. In addition, research on price forecasting contributes to the discovery of certain relationships between the analyzed values. Price fluctuations on the stock exchange seem random from a long-term perspective, however, by observing their behavior, it is possible to find certain patterns that can be used in learning models.

The traditional approach to price prediction uses linear statistical time series models such as ARMA or ARIMA. It should be noted that the linear techniques become suboptimal in the stock market, and non-linear models such as ARCH tend to have lower prediction errors. Observing this relationship, researchers turned their attention to the methods and the techniques that use big data tools and machine learning algorithms to predict prices in financial markets [Heo and Yang, 2016]. Because of the development of computer computing power, it was possible to extend statistical analyzes with new tools from the machine learning portfolio including the Random Forest (RF) and the Support Vector Machine (SVM) algorithms. These algorithms try to find patterns in the data to solve a given problem. Big data and machine learning techniques are also the basis for algorithmic and high-frequency trading routines used by financial institutions

Due to Russia's invasion of Ukraine in 2022 and the restriction of imports of oil and gas from Russia by European countries, the discussion on obtaining alternative energy sources that could solve the problem of potentially more expensive energy carriers has begun anew. Investors also began to look for opportunities to earn money on the fact that in the short term many European economies will have to switch to other alternative energy sources. One of the first thoughts of investors about what energy asset could rapidly increase its price was uranium. These opinions are confirmed by the price of the uranium futures contracts, which at the beginning of 2022 amounted to USD 44, to reach USD 60 already on March 10 (shortly after the start of the war in Ukraine). The increase in the uranium price is related to the limitation of its supply due to the embargo imposed by the USA on the Russian company Rosatom, which is a leading supplier of this raw material [Rosik, 2022]. In addition, markets began to discount the future increased demand for this raw material by European economies, which use it to produce electricity in nuclear power plants. What is more, many countries, such as Poland, declared their willingness to build new nuclear power plants in order to increase energy independence from Russia.

Due to the fact that the prediction of share prices of the uranium companies sector is becoming an object of interest for retail and institutional investors, and there are few articles in the literature on the use of machine learning methods in price forecasting of this sector, the aim of this article was preparing prediction for five periods using the RF and SVM algorithms and comparing the errors of these predictions in order to determine which of the analyzed algorithms is characterized by a lower predictions bias.

Literature review

There are several works which focus on applying machine learning algorithms for solving prediction problems. The random forest algorithm was used by Mei et al. [2014] for forecasting prices on the New York electricity market in real time, which testified to the speed of the learning algorithms. The RF algorithm was also used by Herrera et al. [2010] as a predictive model in the problem of forecasting the hourly water demand of cities. For their analysis, the researchers used two algorithms to forecast the closing price of water contracts: the ANN (Artificial Neural Network) and the RF algorithm. Appling financial data, the researchers created new variables such as opening price, closing price, maximum price and minimum price of a sample of companies. The research showed that the use of the new indicators together with the application of the analyzed techniques was crucial in the forecasting process. What is more the RMSE and MAPE errors were at a satisfactory low level. The research presented by Vijh et al. [2020] also focused on the use of AAN and RF algorithms with additional preparation of new variables from existing financial data. The researchers compared the two models in predicting stock market return from companies from different sectors. Researchers presented that the forecasts prepared using the RF algorithm are within the assumed error level calculated as RMSE, MAPE and MBE. Kumar et al. [2018] used machine learning models such as Random Forest (RF), Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbor (KNN), and Softmax to predict stock market prices. Researchers compared the analyzed methods using several technical indicators based on collected data from a few sources. The results of the study showed that for long time series the best predictive model was the RF algorithm, while for short time series the Naive Bayes algorithm. Another observation made was, as the count of technical indicators was reduced the accuracy of the models decreased. Yuan et al. [2020] used the sliding window method to study the possibility of appling the RF in trend prediction and the selection of model features. The empirical results show that the best performance can be obtained when the RF is applied for both feature selection and stock price trend forecasting. By selecting different stock numbers to build the model, it was also found that the RF model had the highest return when it chosen top 1% of the stocks, achieving a 29.51% annualized return. Results of two teams of researchers Khaidem et al [2016] and Gholamian and Davoodi [2018] proved that the use of the RF algorithm is helpful in minimizing risk when investing in the stock market. The results of the accuracy of the RG algorithm for forecasting stock prices in the quoted study amounted to 64%.

Many articles focus on using the SVM algorithm for both stock price prediction and trend analysis. For example**,** Lin et al. [2013] used the SVM algorithm to predict the future trend of companies on the stock exchange. Das and Padhy [2012] compared two different algorithms found that the best results in terms of forecast accuracy were obtained by the SVM method. Heo and Yang [2016] conducted a study in which the SVM algorithm was used to predict stock price fluctuations. In this study, it was assessed whether the SVM algorithm can predict price fluctuations and what level of forecast error the algorithm will for forecasts for longer periods. As a result, it was found that stock price predictability utilizing financial information input with SVM showed superior predictability to expert's predictions, and that predictability decreases as time goes by. Leung et al. [2014] used a combination of the SVM algorithm together with graphs. Researchers used minimum graph cuts as parts of a cutting plane algorithm to solve the optimization problem of the structural SVM. This approach allowed the algorithm to learn a prediction model for a complex graph input with multiple edges per node (representing complex relationships between companies that affect the stock price by using feature vectors that contain fundamental financial information). The accuracy of of the model was 78%, which meant that the model was not overfitted. Research by Yang et al. [2020] focused on applying of the SVM algorithm to analyze the volatility of financial bottoms of listed companies from Hong Kong's Hang Seng Index. The study showed that the SVM algorithm predicted the closing prices from the analyzed index with the lowest error. Kim [2003] showed that the SVM algorithm is able to correctly predict the direction of daily changes in the Korean composite Stock Price Index 200 (KOSPI 200). Comparison of the SVM algorithm and ANN showed that SVM generates a lower error rate than ANN for the analyzed market.

Methods

Random Forest

The Random Forest method is a nonparametric machine learning algorithm which is constructed in such a way as to use multiple decision trees in its learning process and outputs mean prediction of the individual trees. A decision tree can be presented as a hierarchical analysis diagram in which each internal node represents a test function on one independent variable, each branch represents the test outcome and each terminal (leaf) node represents a decision. The operation of the algorithm is that at each node searches the values of the incoming dataset and recognizes a threshold for one predictor variable to divide the dataset such that the homogeneity of dependent variable values in each branch is maximized [Xu et al., 2019]. The results obtained with this method is the averaging of the results obtained from each tree separately. The idea was originally proposed by J.M Bates and C.W.J. Granger in 1969. Researchers showed that combining two separate predictions for airline passengers can lead to a model that presents a lower root mean square error [Bates and Granger, 1969].

In the Random Forest method, each decision tree is trained using a subset of data randomly sampled with replacement from the original training dataset, which increases the robustness against overfitting [Zhong, et al., 2015]. In order to add an additional layer of randomness, instead of using all provided to model variables, only a randomly selected sample of the variables are considered to split nodes of each tree. The reason to add this additional randomness is to decrease the redundancy of predictor variables while increasing the diversity of the trees in a forest [Xu et al., 2019, Lujan-Moreno, 2018].

The Random Forest algorithm can be describe as presented below:

1. From the dataset select k random records from T training dataset.
2. Based on k records, build a decision tree.
3. From the RF algorithm, the number of trees are chosen and the steps 1 and 2 are repeated.
4. In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output) [Polamuri and Srinivas, 2019]

Suppose that there is a training dataset on prices of uranium companies described as *T, k* different training sample sets are drawn from *T*, denoted by *S*, and the training process generates *k* different learners f(*S*). Then, the final prices of uranium companies prediction can be presented as [Hu et al., 2022]:

$$\hat{\theta} = \frac{1}{k}\sum_{k=1}^{k} f(S) \tag{1}$$

The RF algorithm evaluates the importance of provided in the model variables by estimating how the selected variable affects the prediction error when is chosen in the permutation process, while all other variables remain unchanged. Because of the process, it is possible to check whether the variables selected for the model are significant in the forecasting [Archer and Kimes, 2008]. The RF also reports statistical metrics to assess model performance, including R^2 and prediction error values. In this article, the RF algorithm is implemented using Python sklearn ensemble library and prepared to this process special built in the library RandomForestRegressor function.

Support Vector Machine

The SVM is a classifier of supervised machine learning, also known as a support vector network. It was proposed by Vladimir N. Vapnik and A. Y. Chervonenkis in 1963 as an algorithm which can be apply in classification, regression and detection outliers problems but over time his role has changed and now the method is used in many other calculations [Madhu at al., 2019].

The SVMs are linear learning machines which means that a linear function is always used to solve the regression problem. When dealing with nonlinear regression, the SVMs map the data x into a high-dimensional feature space via a nonlinear mapping φ and make linear regression in this space. The function describes the process can be wrote as follows [Xie at al., 2006]:

$$f(x) = w * \varphi(x) + b \tag{2}$$

where φ is called the nonlinear feature space mapped from input space x and y is the estimated value in terms of input data x. Coefficients w and b are estimated by minimizing [Lean at al., 2017]:

$$R_{reg}(C) = C * \frac{1}{N}\sum_{i=1}^{N} L_{\varepsilon}(d_i, y_i) + \frac{1}{2}||w||^2 \tag{3}$$

$$L_\varepsilon(d_i, y_i) = \begin{cases} |d-y| - \varepsilon, |d-y| \geq \varepsilon \\ 0, otherwise \end{cases} \quad (4)$$

where R_{reg} is the regularized risk function, d_i is the actual value in the i th period, and C and ε are parameters that are specified by an user. In Eq. (3), the first part $C * \frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i)$ is the empirical error, measured by the ε-insensitive loss function given by Eq. (3). This loss function allows to enable one to use sparse data points to represent the regression function defined by Eq. (2). The second term, $\frac{1}{2}||w||^2$, is the regularization part, which sets the flatness of the function. C is referred to as the regularized constant and it determines the trade-off between empirical risk and the regularization term. Increasing the value of C will result in relatively higher importance of the empirical risk. ε is called the tube size, and it is equivalent to the approximate accuracy placed on the training data points minimizing [Lean at al., 2017].

In addition to specifying the regularized constant (*C*) and parameter ε functions by the user, the kernel function must also be specified. The kernel function simplifies the use of a mapping. The function is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. The kernel can take many forms such as: linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. [Data Fair, 2023].
In sklearn library there are four kerner functions that can be use in SVM algorithm [scikit-learn, 2023]:

1. linear: $<x, x'>$,
2. polynomial: $(\gamma<x, x'>+r^d$, where d is specified by parameter degree, r by coef0
3. rbf: exp $(-\gamma||x - x'||^2)$, where γ is specified by parameter gamma and must be greater than 0,
4. sigmoid: tanh $(\gamma < x, x' > +r$, where r is specified by coef0.

In this article, the SVM algorithm is implemented using Python sklearn svm library with SVR function prepared for the regression problem.

The data and the data preparation

To achieve the goal set in the article, data on the ETF Global X Uranium for the period from 08/11/2010 to 31/05/2023 was obtained from investing.com. The ETF was selected because the financial instrument seeks to track the Solactive Global Uranium & Nuclear Components index which includes companies worldwide that are engaged in the exploration, mining and/or refining of uranium. The data contains information about the stock such as High, Low, Open, Close, Adjacent close and Volume. According to the approach proposed by Vijh et al. [2020] several well-known technical indicators like the relative strength indicator (RSI), stochastic oscillator (slow, fast, moving average cross-over divergence (MACD), price rate of change (ROC), on balance volume, money flow index (MFI), Williams accumulation and distribution (WAD), and the 50-day and 200-day moving averages, calculated from daily data, are used as features in the prediction models. The number of observations used to estimate the models varies and depends upon the calculation of the technical indicators (200 days are omitted due to the calculation of the 200-day moving average) and forecast periods (between 1 and 5 observations are omitted depending upon the forecast horizon). For a 5-day forecast horizon, there are 2979 observations. For the analysis, 80% of the data was used for training and 20% used for testing. The training data set consists of 2383 observations (80% of the data) and the testing data set contains 596 observations (20% of the data). Because data used for the SVM algorithm needs to be standardized, function StandardScaler from Sklearn python library was used. Both the RF and the SVM algorithm must have properly tuned hyperparameters. Fine-tuning of hyperparameters can be done by trial and error or by using the appropriate GridSearchCV functions from the sklearn library. In the presented article, the GridSearchCV function was used, to check which a set of candidate hyperparameters is optimal. In conducting sensitivity analysis, training control for the RF and the SVM was handled with 10-fold cross validation with 10 repeats.

The hyper-parameter tuning tested the following parameter combinations for the RF algorithm: bootstrapping as the sample selection method; a number of trees varying by ten between 200 and 1000; 2, 5, or 10 as the minimum number of samples required to split a node; 1, 2, 4, or 5 as the minimum number of samples required at each leaf node; all features or subset of features allowed for tree consideration; and no limit on the maximum tree depth. The best-fit hyperparameters for RF turned out to be the following set of hyperparameters:

- max futures: 4
- min samples leaf: 1
- min samples split: 2
- number of decision trees: 500.

The number of decision trees is sufficient to be able to determine that the model will not be overtrained. In choosing the correct set of the hyperparameters it is crucial to remember that a very large number of trees does not lead to overfitting, but a small number of trees results in high test error.
The following set of hyperparameters were tested for the SVM algorithm: regularized constant from the candidate set: 0.1, 1, 10, 100, 1000, ε parameter from the candidate set:1, 0.1, 0.01,, 0.05 0.001, 0.0001, kernel from the candidate set: linear, rbf and sigmoid. The best-fit hyperparameters for SVM algorithm are presented below:

- regularized constant: 100,
- ε: 0.05,
- kernel: rbf.

This selected parameter was set for the Random Forest and the Support Vector Machine estimation in this study.

Results

As presented in the previous section, the data used in the study was the ETF Global X Uranium (URA) prices from 08/11/2010 to 31/05/2023. The figure 1 shows the daily prices of the analyzed ETF over the period under the study.

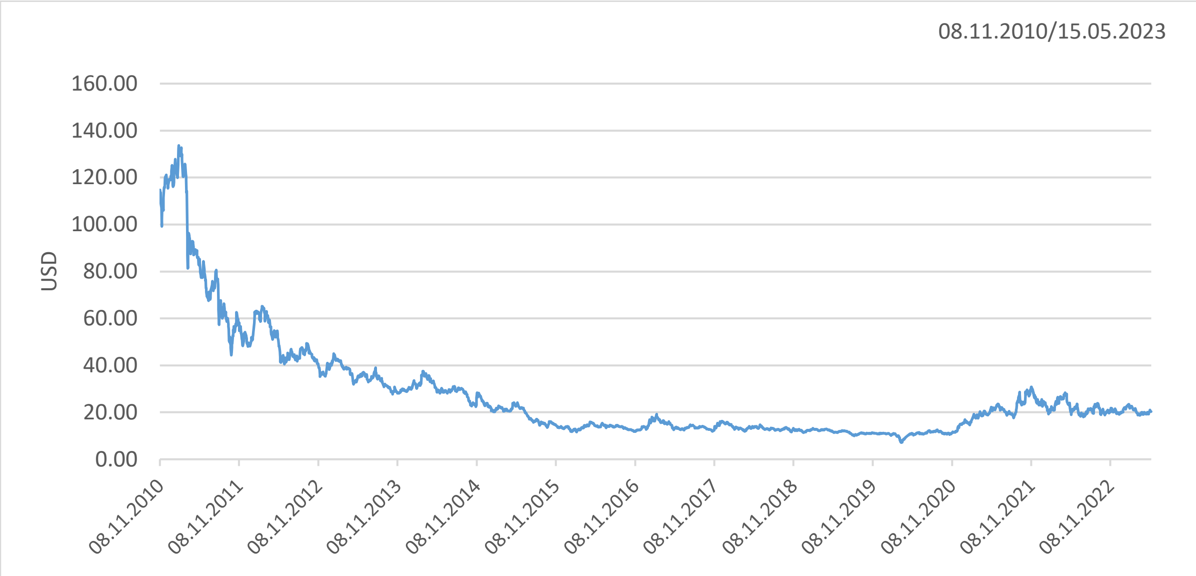

Figure 1. This figure shows ETF Global X Uranium prices across time. Data are sourced from investing.com.

The figure shows a peak at the beginning of the analyzed period. The reason for the high ETF prices in that moment was debt issues in Europe and the United States and related concerns of inflation but also because real interest rates turned negative. The disaster at the nuclear power plant in Fukushima, which took place on April 11, 2011, significantly disrupted the upward trend by increasing anti-nuclear sentiments. Many countries in the world, such as Germany, France, Belgium and Japan started to make energy transformations, turning away from nuclear power in favor of oil, gas and renewable energy sources. The decline in ETF prices continued until 2015, when the ETF price was at the level of 12-16 USD. The sideways trend in ETF prices continued until 2021, when the prices began to rise. That trend was strengthened by the start of the war in Ukraine and the growing risk of blackouts in a number of countries around the world. In addition, ETF prices were also affected by the EU regulation on including nuclear energy in the so-called EU taxonomy, i.e. a list of "green" energy sources that will be able to obtain funding from the European Union [Suder, 2023].

Descriptive statistics of the daily closing prices of the Global X Uranium ETF are presented in Table 1.

Table 1. Descriptive statistics of the daily closing prices (in USD) from 08/11/2010 to 31/05/2023

Median	27.11
Mean	19,67
Min	7,09
Max	133,68
Std.dev	22,44
Skewness	2,49

Kurtosis	6,81

Source: own study based on data from investing.com

Analyzing the data in Table 1, it can be noticed that the average ETF price was 19.67 USD with a standard deviation of 22.44 USD, and the median was 27.11 USD. The minimum value oscillated around 7.09 USD and the maximum 133.68 USD. Kurtosis at the level of 6.81 means that the distribution has the so-called "fat tails", i.e. the intensity of extreme values is greater than in the normal distribution. A skewness 2.48 indicates a right-sided distribution.

The financial indicators discussed in the previous session were created. On their basis, prices of the URA were predicted for the next five periods (one full financial week) using fine-tuned the RF and the SVM algorithms. In order to be able to compare the results of machine learning algorithms, it was decided to prepare an additional set of predictions using the ARIMA model. Before preparing the forecasts using the ARIMA method, the Dickey-Fuller test (ADF) was applied to check the stationarity of the time series. This test showed that the recalculation of first-order differences needs to be calculated to obtain stationarity. Then, using the Akaike criterion, the best set of parameters were checked. The final ARIMA model which was used to prepared the predictions was ARIMA(2,1,1).

Table 2 presents the adjustment of the analyzed algorithms to the historical data. Four measures were used for the analysis: R^2, MAE, MSE, RMSE.

Table 2 Fitting the models to the historical data

	ARIMA (2,1,1)	RF	SVM
R^2	0.993	0.998	0.994
MAE	0.364	0.197	0.349
MSE	0.339	0.128	0.335
RMSE	0.583	0.358	0.579

Source: own study based on data from investing.com

Analyzing table 2 it can be noticed that R^2 is at a very high level of 0.99. Such a good fit of all the analyzed models to the data results from the fact that only ETF prices and price-derived indices were used for modelling. It is worth noting that the MAE, MSE and RMSE errors for the ARIMA model and the SVR algorithm are at a similar level, while the RF algorithm is characterized by a lower level of forecast bias.

To forecast the future prices the approach proposed by Sadorsky [2021] was used. The forecast variable is Global X Uranium ETF price and the features are technical indicators, some of which (like the MA200) embody a lot of historical data about stock prices that helps to mitigate the residual serial correlation. To investigate this issue further, a time series cross validation analysis is conducted where the first 80% of the data are used to fit a RF and SVM algorithms and price predictions are made. Then, the estimation sample is increased by one observation and the model re-fit and a new set of forecasts produced. This recursive approach is used until the end of the data set is reached. This approach representative of what an investor actually does in practice. The table 3 presents five-period forecasts using the RF, the SVM and the ARIMA together with actual price realizations and the forecast error. The figure 2 presents the forecasts together with the real prices.

Table 3 Comparison of forecasts for 5 periods with actual price executions

Date	Real URA price in USD	RF prediction in USD	Diff %	ARIMA(2,1,1) prediction in USD	Diff %	SVM prediction in USD	Diff %
17.05.2023	20,02	20,22	0,99%	19,77	-1,25%	19,90	-0,62%
18.05.2023	20,06	20,27	1,03%	19,74	-1,60%	20,007	-0,26%
19.05.2023	20,11	20,37	1,30%	19,73	-1,89%	20,060	-0,25%
22.05.2023	20,70	20,44	-1,26%	19,71	-4,78%	20,088	-2,96%
23.05.2023	20,44	20,43	-0,03%	19,70	-3,62%	20,073	-1,80%

Source: own study based on data from investing.com

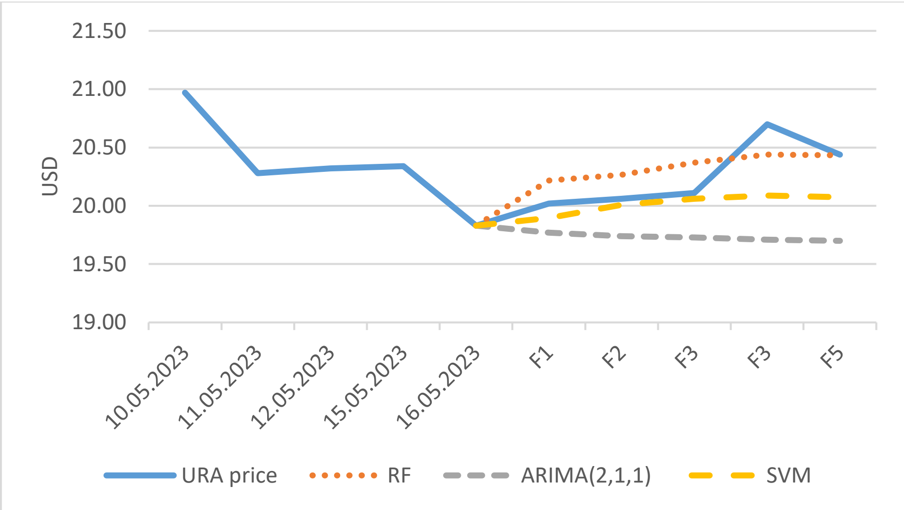

Figure 2. This figure shows ETF Global X Uranium realization of prices with the predictions of the models.

Analyzing the above results, it can be seen that the RF algorithm was characterized by a stable forecast error to the actual execution of the ETF prices throughout the period. Interestingly, the algorithm overestimated the first three forecasts upwards, while the next two downwards. The change from the direction of revaluation was associated with a strong increase in the real ETF price. The ARIMA model and the SVM algorithm generated underestimated predictions, with the predictions of the SVM algorithm being more accurate than those generated by the ARIMA. Both the ARIMA and the SVM had an issue with the moment of rapid price increase on May 22, 2020, while the SVM algorithm reacted to the price change faster than the ARIMA model, finally generating the last forecast with a smaller error than the classic model. It can be assumed that in the case of a stable price exchange rate, the SVM model would be characterized by the lowest forecast error among the analyzed models.

Discussion and Summary

The aim of this article was preparing forecasts using the RF and the SVM algorithms and comparing the errors of obtained predictions in order to determine which of the analyzed method is characterized by a lower predictions bias.

The results of the experiment showed the legitimacy of using the RF and the SVM algorithms in forecasting ETF prices of uranium companies. Both the SVM and the RF algorithm can eliminate some of the disadvantages of conventional methods, e.g., local minima and over-fitting problems, in terms of empirical risk minimization (ERM) principle, and thus obtain more stable and robust generalization results, relative to conventional methods [Tay and Cao, 2001]. However, it should be remembered that for machine learning properly selected hyperparameters must be done. In the case of the SVM algorithm, the regularization constant C, the $\boldsymbol{\varepsilon}$ and the kernel should be optimized, while for the RF algorithm, the attentione should be paid on the number of variables taken into the model, tree division, and the number of decision trees.

The conducted research presented that the predictions calculated by the RF and SVM algorithms are less biased than the predictions generated by the ARIMA model. It is worth noting that the SVM forecasts were the most accurate in the first three periods, while when the price rose, the model reacted with a delay. In the case of the RF algorithm, the forecast error remained at a similar level almost throughout the forecast period. The exception may be the fifth period, for which the forecast error was insignificant compared to the actual value, but this was due to the fact that the model, on average, overestimated the previous forecasts and at the time of the sudden price increase, the forecast values were closer to their actual realization.

However, one should remember about some important elements of the price of uranium companies, which cannot be examined by analyzing only the price of a financial instrument and the indicators based on this price. Many irregular events also have a significant impact on the price of uranium companies (an example of the disaster at the nuclear power plant in Fukushima in 2011), but they are extremely difficult to detect and use in the analysis of price volatility and prediction. Certainly, a further direction of research may be to add more variables to the algorithms that would be able to smooth out irregular fluctuations or to use combined models in which both classical statistical models and machine learning algorithms are combined into one model.

REFERENCES

Heo, J., Yong Yang, J., 2016. *Stock Price Prediction Based on Financial Statements Using SVM*, International Journal of Hybrid Information Technology, Vol 9., No.2, pp. 57-66.

Rosik, P., 2022. *Uran jest najdroższy od 11 lat. Jego cena może nadal rosnąć, bo największy producent z Rosji może zostać objęty sankcjami*, Strefa Inwestorów, online: https://strefainwestorow.pl/artykuly/surowce/20220315/uran-hossa-wojna-ukraina-rosja [13.06.2022].

Mei, J., He, D., Harley R., Habetler, T., Qu, G., 2014. *A random forest method for real-time price forecasting in New York electricity market,* 014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, USA, pp. 1-5, doi: 10.1109/PESGM.2014.6939932.

Herrera, M., Torgo, L., Izquierdo, J., Pérez-García, R., 2010. *Predictive models for forecasting hourly urban water demand,* No. 387, pp. 141–150.

Vijh, M., Chandola, D., Tikkiwal, A., V., Kumar, A., 2020. *Stock Closing Price Prediction using Machine Learning Techniques,* Procedia Computer Science, Vol. 167, pp. 599-606.

Kumar, M., Thenmozhi, M., 2006. *Forecasting stock index movement: A comparison of support vector machines and random forest*, In Indian institute of capital markets 9th capital markets conference paper.

Yuan, X., Yuan, J., Jiang, T., Ain, U,. Q,. 2020. *Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market*, IEEE Access, Vol. 8, pp. 22672 – 22685.

Khaidem, L,. Saha, S., Basak, S., Kar, S., Dey, S,. 2016. *Predicting the direction of stock market prices using random forest.*

Gholamian, E,. Davoodi, S.M.R., 2018. Predicting the direction of stock market prices using random forest.

Lin, Y., Guo, H., Hu, J., 2013. An SVM-based approach for stock market trend prediction. Paper presented at the The 2013 international joint conference on neural networks (IJCNN).

Prasad Das, S., Padhy, S., 2012. Support Vector Machines for Prediction of Futures Prices in Indian Stock Market. International Journal of Computer Applications. Vol. 41, pp. 22-26.

Leung, C., K-S., MacKinnon, K., R., Wang, Y., 2014. A machine learning approach for stock price prediction. In Proceedings of the 18th International Database Engineering & Applications Symposium (IDEAS '14). Association for Computing Machinery, New York, NY, USA, pp. 274–277.

Kim, K.-J., 2003, Financial time series forecasting using support vector machines, Neurocomputing, vol. 55, pp. 307-319.

Xu Z., Lian J., Bin L., Hua K., Xu K., Chan HY., 2019. Water Price Prediction for Increasing Market Efficiency Using Random Forest Regression: A Case Study in the Western United States, Water, Vol.11, no. 2, 228.

Bates, J. M., Granger, C. W. J. 1969. The Combination of Forecasts, Vol. 20, no. 4, pp.451–468.

Zhong, S., Xie, X., Lin, L., 2015. Two-layer random forests model for case reuse in case-based reasoning, Expert Systems with Applications, Vol. 42, no. 24, pp. 9412-9425,

Lujan-Moreno, G.A.; Howard, P.R.; Rojas, O.G. Montgomery, D.C., 2018. Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. Expert Syst, Vol. 109, pp. 195–205.

Polamuri, R., S., Srinivas, K., Mohan, K., A., 2019. Stock Market Prices Prediction using Random Forest and Extra Tree Regression. International Journal of Recent Technology and Engineering, vol. 8, pp. 1224 – 1228.

Hu, L., Chun, Y., Griffith, D. A., 2022. Incorporating spatial autocorrelation into house sale price prediction using random forest model. Transactions in GIS, Vol. 26, pp. 2123– 2144.

Archer, K. J., & Kimes, R. V., 2008. Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis, Vol. 52, no. 4, pp. 2249–2260.

Madhu, B., Rahman, M., Mukherjee, A., Islam, M.,Roy, R., Ali, L., 2021. A Comparative Study of Support Vector Machine and Artificial Neural Network for Option Price Prediction. Journal of Computer and Communications, Vol. 9, pp. 78-91.

Xie, W., Yu, L., Xu, S., Wang, S., 2006. A new method for crude oil price forecasting based on support vector machines. In Proceedings of the 6th international conference on Computational Science - Volume Part IV (ICCS'06), Vol. Part IV. Springer-Verlag, Berlin, Heidelberg, pp. 444–451.

Yu, L., Zhang, X., Wang, S., 2017. Assessing Potentiality of Support Vector Machine Method in Crude Oil Price Forecasting. Eurasia Journal of Mathematics, Science and Technology Education. Vol. 13, no. 12, pp. 7893-7904.

Data Fair, 2023. Kernel Functions-Introduction to SVM Kernel & Examples, online: https://data-flair.training/blogs/svm-kernel-functions/ [13.06.2023].

Scikit-learn, 2023. Support Vector Machines, online: https://scikit-learn.org/stable/modules/svm.html#kernel-functions [13.06.2023].

Suder, J., 2023. Wzrost zapotrzebowania na energię jądrową zwiększa globalny popyt na uran, Obserwator finansowy.pl, online: https://www.obserwatorfinansowy.pl/tematyka/makroekonomia/trendy-gospodarcze/wzrost-zapotrzebowania-na-energie-jadrowa-zwieksza-globalny-popyt-na-uran/ [13.06.2023].

Sadorsky, P., 2021. Predicting Gold and Silver Price Direction Using Tree-Based Classifiers. J. Risk Financial Manag. Vol.14, 198. https://doi.org/10.3390/jrfm14050198

Tay, F. E. H., Cao, L. J., 2001. Applications of support vector machines in financial time series forecasting. Omega, Vol. 29, pp. 309-317.