Yayına Geliş Tarihi:15/02/2022
Yayına Kabul Tarihi:12/06/2022
Online Yayın Tarihi:28/07/2022

Meriç Uluslararası Sosyal ve Stratejik Araştırmalar Dergisi
Cilt:6, Sayı:16, Yıl:2022, Sayfa:124-143
ISSN: 2587-2206

NASDAQ 100 INDEX ESTIMATE WITH MACHINE LEARNING ALGORITHMS

Zeynep ŞENGÜL[1]

Abstract

Investors want to know about the future to ensure maximum profit and minor losses before making an investment decision. Today, high-level fluctuations in the stock market are influenced not only by the economy, but also by external factors, resulting in sudden ups and downs. In this study, Linear, Polynesic, Sigmoid and Radial based Support Vector regression, Random Forest regression and K-nearest neighbors regression models were implemented using the daily closing values of the NASDAQ 100 index, which included 2016-2021 from www.investing.com, for the next day price estimate. Given the MSE, RMSE, MPE, and R^2 values set for model performance evaluation, the best estimate was given by Radial based Support Vector Regression.

The study concludes that the regressions of Radial based Support Vectors, Random Forest and the k-Nearest Neighbors can be preferred from machine learning algorithms used in practice instead of traditional econometric methods for future prediction in financial series.

Keywords: *NASDAQ 100 Index, Support Vector Regression, Random Forest Regression, K - Nearest Neighbors Regression*

[1] Trakya Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Yüksek Lisans Öğrencisi, zzeynepsengul@gmail.com, ORCID: 0000-0002-0461-6203

Atıf/Citation:Şengül, Z. (2022). Nasdaq 100 Index Estimate With Machine Learning Algorithms. *Meriç Uluslararası Sosyal ve Stratejik Araştırmalar Dergisi*, 6(16), 124-143.

MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE NASDAQ 100 İNDEKS TAHMİNİ

Özet

Yatırımcılar, bir yatırım kararı vermeden önce maksimum kar sağlamak ve minumum zarara uğramak için gelecek hakkında bilgi sahibi olmak ismektedirler. Bugün borsadaki yüksek oynaklıklar sadece ekonomiden değil, dış faktörlerden de etkilenerek ani iniş ve çıkışlara sebep olmaktadır. Bu çalışmada, www.investing.com'dan elde edilen 2016-2021 yılları arasını kapsayan NASDAQ 100 endeksine ait günlük kapanış değerleri kullanılarak bir sonraki gün fiyat tahmini için makine öğrenmesi algortimaları olan Doğrusal, Polinominal, Sigmoid ve Radyal tabanlı Destek Vektör regresyonu, Rastgele Orman regresyonu ve K-en yakın komşular regresyon modelleri uygulanmıştır. Model performans değerlendirme için belirlenen MSE, RMSE, MPE ve R^2 *değerleri dikkate alındığında en iyi tahmini Radyal tabanlı Destek Vektör Regresyonunun verdiği belirlenmiştir.*

Çalışma kapsamında finansal serilerde gelecek tahmini için geleneksel ekonometrik yöntemler yerine uygulamada kullanılan makine öğrenmesi algoritmalarından Radyal tabanlı destek vektör, rastgele orman ve k en yakın komşular regresyonunlarının tercih edilebileceği sonucuna varılmaktadır.

Anahtar Kelimler: *NASDAQ 100 Endeksi, Destek Vektör Regresyonu, Rastgele Orman Regresyonu, K-En Yakın Komşular Regresyonu*

1. INTRODUCTION

It is difficult for all investors to make an investment decision in the globalized world today. This is due to continuous change and inaccurate situations. This process has a negative impact on investors' decision on investment.

The stock market's future price is an important issue of their predictability. Investors want to minimize the risk in their investments in stocks and maximize the profit. The stock market indexes are affected by any positive or negative events in the world, and have high uncertainty and high play. Because stocks are affected by many factors that are economic or not. It is not possible to make a definitive prediction because of this situation. Historical values are being looked at for the forecast of future prices of stock shares. The reason they're looking at past values is because they predict how stock market movements will be affected in an unusual situation.

These predictive methods have been partially successful for the future forecast due to the moving structures of stocks. Machine learning algorithms developed outside of traditional methods are now being preferred for stocks estimates.

This study is not at the level of daily data obtained between 2016 and 2021 of the NASDAQ index. The data was obtained from www.investing.com and the scikit-learn library was used by writing the application in python. So with the return of the series, it was analyzed using the Support vector machine, Random forest and the K-nearest neighbors algorithm. Error values are taken into account as a performance assessment.

2. LITERATURE

When the relevant literatüre is examined, it appears that machine learning algorithms and many analyzes have been carried out anda re being used in the time series.

Meesad and Rasel (2013) estimated their work using the Support Vector Regression for market forecast. As time series data, the Dhaka Stock Exchange company has been compared using ACI limited company data by performing a Support Vector Regression estimate. The Support Vector Regression is said to be a strong predictor.

Filiz et al. (2017), in their study, they analyzed the BIST-50 index using machine learning methods and artificial neural network. Arguments for dependent variable BIST-50; DAX, S&P 500, BISTSINAI, BISTMALI, FTSE, BISTBANK, Raw oil Price, Euro/Dollar Parity, interest rates and GOLDINDEX are used. The analysis uses the k-nearest neighbors, Naive Bayes, C4.5 and artificial neural Network models. The analysis results in the best performance showing the C4.5 algorithm.

Polamuri et al. (2019), they wanted to predict the market index behavior with a new method for analysis, and they used Linear Regression, Support Vector Regression, Decision Tree Regression, Random Forest Regression, and Extra Tree Regression from machine learning algorithms. They said the best predictor was a decision tree and Random Forest Regression.

Alhnaity and Abbod (2020) analyzed their work to find the closing value of the London Stock Exchange index, S&P 500 and Tokyo Stock Exchange indexes to the future market. In the analysis, Support Vector developed a model based on the methods of Regression, Ensemble ampiric Mode Parsing, genetic Algorithm, Reverse spread neural Network, and Weight Average analysis. They said that the new approach that was created by predicting and comparing the three-stage new model that they created after performing a single prediction with the analysis methods they used in the model they developed and those models was more accurate.

Vijh et al. (2020) the next day, the share shares used in the study perform analysis using the Artificial Neural Network and Rassal Forest algorithms to estimate the stock closing values. As input, the stock market opening price values, the highest values in the day, the lowest values in the day, and the stock exchange closing values are used. Average Square Error (RMSE) and Average Absolute percentage Error (MAPE) values are compared to analyze model performance values of estimates. Because the calculation results of the error values are low values, it has been found that the Artificial Neural Networks and Rassal Forest methods are appropriate methods for estimating stock closing prices.

Arslankaya and Toprak (2021) have used Polyinom Regression and Rassal Forest Regression, RNN, and Long-term Memory (LSTM) methods, which repeat deep learning methods, through machine learning methods to make stock price estimates in their work. The most effective result is the Random Forest Regression model, and the worst result is the Polinom Regression model.

3. MATERIAL and METHOD

3.1.Support Vector Machines

The support vector machine is a powerful machine learning algorithm developed by Vapnik. The support vector machine is an algorithm used to first perform classification (Vapnik, 1995). When the support vector machine is initially used as an algorithm capable of only two classification, this algorithm can be developed and divided into two and more classes and used in regression problems.

Support vector regression is based on minimizing risk. This makes it preferable in analytics. Support vector regression has functions for both linear and nonlinear data. Core functions are used for nonlinear models.

The support vector machine (SVM) has a linear separation function. This is used in data that can be separated linearly. The goal is to find the best classification plane for data. Given that we have a linear data set, an infinite number of hyperplanes can be drawn to best divide this data set into two classes. The maximum distance between these two classes is called the margin.

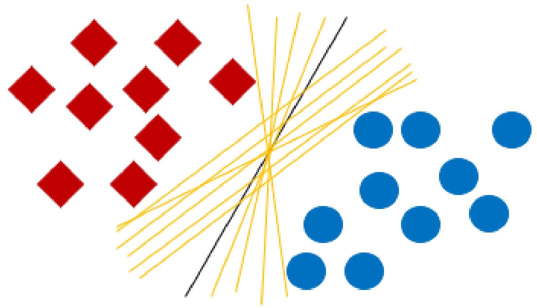


Figure 1.1: Classification separated linearally

The data contained in Figure 1.1 is linear. As shown in the figure, there are infinite hyperplanes separating these two classes. There is at least one error and one hyperplane that successfully separating these two classes.

Assuming that the data set is n elamen,

$B = \{x_k, y_k\},\ k = 1,2,3,\ldots,n\ \ ,y_k\ \in \{-1, 1\}\ \ \ , x_2 \in \mathrm{R}$

The x_k input vector in the above function indicates that the y_k samples have fallen into the {-1,1} class.

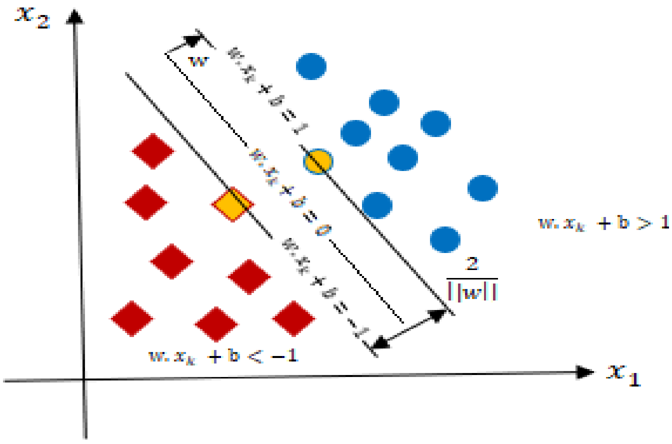


Figure 1.2: Suitable hyperplane and support vector machines

When the graph in Figure 1.2 is examined, the distance between $w.x_k + b = 1$ and $w.x_k + b = -1$ is $2/||w||$. The support vector machine is trying to make this distance too large. To obtain the anthems of my Maxima, the expression $||w||$ must be as small as possible.

$$w.x_k + b = 0 \tag{1.1}$$

In the function in Equality 1.1, the weight vector w represents the threshold value b. The largest margin limit of samples in the dataset to find the appropriate hyperplane,

$$w.x_k + b \geq 1$$

$$w.x_k + b \leq -1 \tag{1.2}$$

The equation is expressed by the functions in 1.2. These functions indicate that the observations on the appropriate hyperplane are. $w.x_k + b \geq 1$, as shown in the graph in figure 1.2, and the observations that fall below the appropriate hyperplane belong to class $w.x_k + b \leq -1$.

For nonlinear data, the error rate may be higher when we split the data set linearly. Nonlinear support vector machine is used to minimize error rate. The nonlinear support vector machine algorithm takes data into high-dimensional space where classification can be performed.

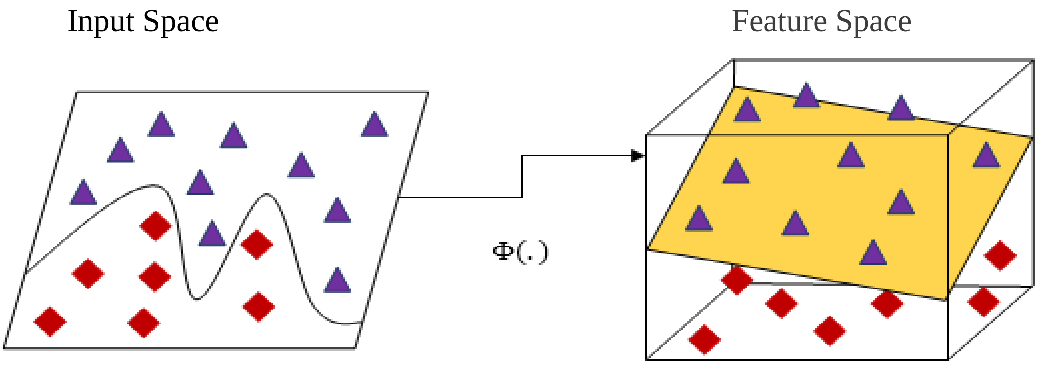


Figure 1.3: Example of two-class data that cannot be separated linearally

When Figure 1.3 is examined, data that cannot be separated linearly can be separated in high-dimensional property space. Thus, data that cannot be separated linearly can be separated with less errors than the linear-separable SVM algorithm and select the appropriate hyperplane without deviation. The nonlinear SVM equation is located in equation 1.3.

$$w.\phi(x) + b = 0 \tag{1.3}$$

Given the equation in 1.3, the equations of the two classes are as shown in the equation in 1.4:

$$w.\phi(x) + b = +1$$

$$w.\phi(x) + b = -1 \tag{1.4}$$

The core function is used because moving data to a higher size on the support vector machine makes it difficult to multiply all values individually. The core function carries data into high-dimensional space. The $\Phi(x_i).\Phi(x_j)$ internal multiplication of the data being transported into space affects the training algorithm. Core function,

$$K(x_{i,}x_j) = \Phi\,(x_i).\Phi(x_j) \tag{1.5}$$

expressed.

As a result, the formula to be used to determine the appropriate hyperplane,

$$f(x) = \left(\sum \alpha_j y_j \Phi\,(x_i).\,\Phi(x_j)\right) + b \qquad (1.6)$$

SVM is used for initial classification. The algorithm is developed with the diversity of problems and performs regression predictions. The support vector regression (SVR) functions like a SVM. The difference between the regression and the classification is that the hyperplane range remains maximum when classifying the data, while the regression problem remains maximum data in the hyperplane range.

The basis of the linear regression model is assumed that there is a linear relationship between the x and y variables, and that the y variable increases as x increases. Assuming that our data meets the conditions of equity in 1.7,

$$(y_{1,}x_1), \ldots\ldots\ldots, (y_{k,}x_k), \;\; x \in R^k, \; y \in R \qquad k = 1,2,3, \ldots\ldots k$$

$$f(x) = w.x_k \; + \; b \qquad (1.7)$$

In the equation x_k represents the input vector in k, the output value y_k, the w weight vector, b deviation. The purpose of the linear machine learning regression is to obtain the function that calculates y_k from x_k.

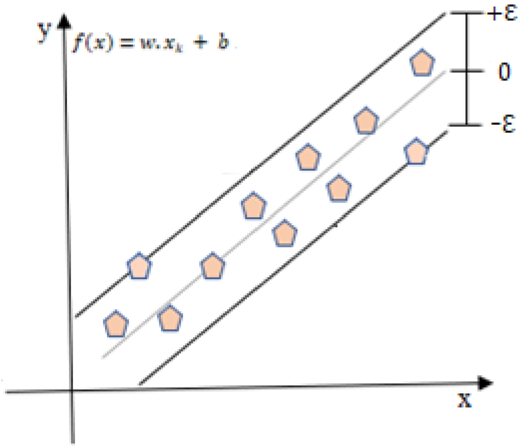


Figure 1.4: Linear support vector function

When the graph in Figure 1.4 is examined, all data is found between the hyperplane. But for all data, it does not apply.

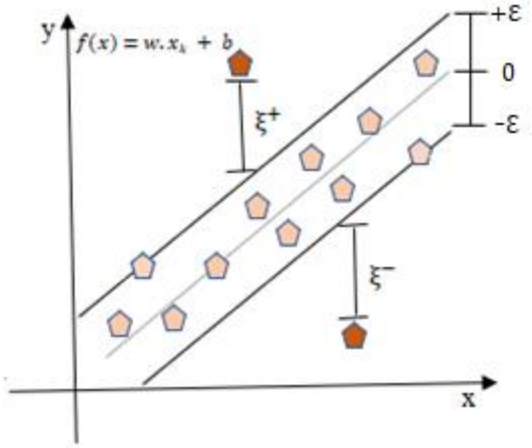


Figure 1.5: Linear support vector regression flexibility variable (ξ) and fault tolerance (ε)

When the graph in Figure 1.5 is examined, it is also shown that data is present between the hyperplane. Data other than hyper-correction indicates that there is no function for them. Not all data is between the hyperplane, resulting in deviations as a result of an estimate. To prevent this situation, the flexibility variable (ξ) is added. The corresponding equation is given in the equation 1.8:

When the graph in Figure 1.5 is examined, it is also shown that data is present between the hyperplane. Data other than hyper-correction indicates that there is no function for them. Not all data is between the hyperplane, resulting in deviations as a result of an estimate. To prevent this situation, the flexibility variable (ξ) is added. The corresponding equation is given in the equation 1.8:

$$
\begin{aligned}
& y_k - w.x_k - b \leq \varepsilon + \xi^{+} \\
& w.x_k + b - y_k \leq \varepsilon + \xi^{-} \qquad (1.8) \\
& \xi^{+}, \xi^{-} \geq 0, \qquad k = 1,2,\dots,N
\end{aligned}
$$

When using a linear model in data that can be separated linearly, in data that cannot be separated linearly, data is moved to higher-dimensional space. This is how the nonlinear SVM works. The nonlinear core function is used instead of the internal multiplication to perform the nonlinear regression estimate. Nonlinear regression prediction function:

$$f(x) = \sum_{i=1}^{l}(\alpha_i^+ - \alpha_i^-)K(x_i, x) + b \quad (1.9)$$

is expressed.

In the formula, $K(x_i, x)$ refers to the Kernel function. Equals x_i and x_j's internal multiplier in high-size input space (Tay and Cao, 2001).

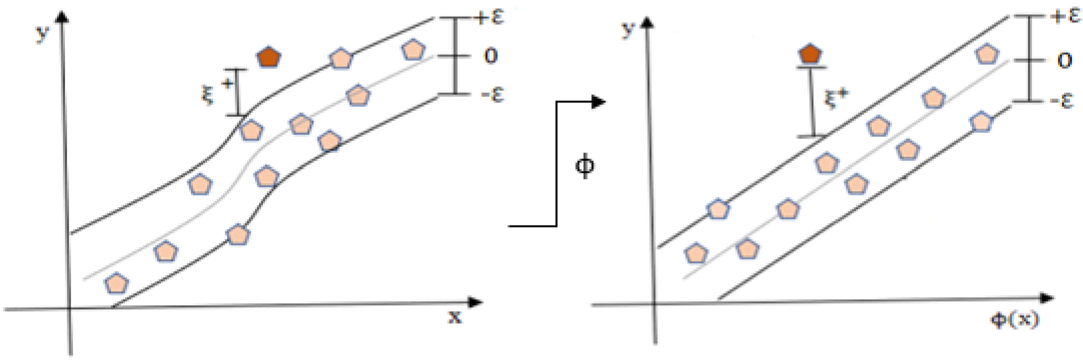


Figure 1.7: Support to Linear Vector Regression flexibility variable (ξ) and fault tolerance (ε)

The core approach is used in regression and classification problems. It provides performance-effective results due to flexibility in the DVM.

The most used core functions are linear core function, polynomial core function, radial-based core function, and sigmoid core function. This is why these four different models of the Support Vector Regression are used in the study.

The linear core function is the simplest core function. The linear core function is located in the equation in 1.10.

$$K(x, y) = x.y \quad (1.10)$$

In the equation in 1.10, x.y creates an internal multiplication.

The polynomial core function is used in nonlinear models, not static.

$$K(x, y) = (1 + x.y)^d \quad (1.11)$$

In the equation in 1.11, 1 refers to the constant term. d refers to the degree of polynomy.

The gamma (γ) parameter, also included in the radial-based core function, can be set. The correct value of this parameter plays an important role in the estimation algorithm. In Equality 1.12, the equation is located.

$$K(x, y) = (e)^{-\gamma||x-y||^2}, \gamma > 0 \quad (1.12)$$

The sigmoid function is a nonlinear SVR model. It is expressed in the relevant study where it performs well.

$$K(x, y) = \tan(\alpha x.y + c) \quad (1.13)$$

In the equation of the sigmoid core function in 1.13, the α slope refers to the term c-cut constant.

3.2. Random Forest

The Random forest (RF) algorithm is the algorithm developed by Breiman (Breiman, 2001). As the RF algorithm is called, it consists of a lot of decision trees coming together, and each tree picks random data. After each tree completes its data process, the forest that occurs begins to vote by starting to classify. The class that gets the most votes gives the RF forecast.

In the RF algorithm, it has two regulating parameters, the number of rassal selected predicters (m) and the number of trees (J) occurring in the forest(Cutler et al. 2012).

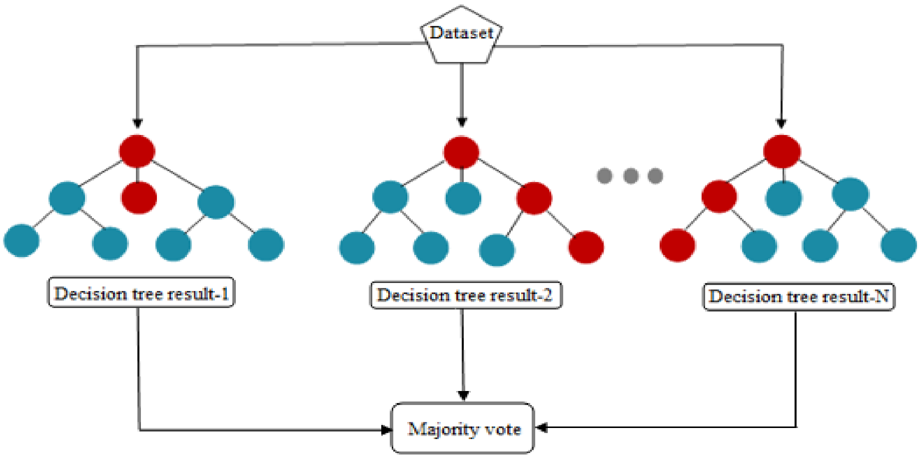


Figure 2.1: Random Forest

The decision tree algorithm shows the tendency to overlearn. The RF algorithm can overcome the problem of overlearning. Unlike the RF algorithm classification problem, it can also be used as a solution to regression problems due to problem variations.

$$E_{XY}(L(Y, f(X))) \tag{2.1}$$

The closer $f(X)$, to Y given in the loss function given in Equality 2.1, the smaller the prediction error. In the RF regression, the lost function is calculated by taking the error frame of the loss:

$$L(Y, f(X)) = (Y - f(X))^2 \tag{2.2}$$

3.3. k - Nearest Neighbors

k - Nearest Neighbor (k-NN) algorithm is used to solve classification and regression problems. This algorithm is one of the controlled learning methods, and its working principle is based on learning data. The data that is not classified is assigned to the most appropriate class based on the calculated distance compared to the data held in the training set (Mitchell, 1997).

The number of neighbors (k), distance measurement and weighting parameters that affect the operating performance of the k-NN algorithm.

The number of neighbors (k) has different characteristics depending on the values received. The class can be created by giving different values from k value 1 to the number of data. When assigning to the group with the closest neighbor in case k=1, the k's greater than the number of data can reduce the performance value by assigning non-similar data to the same group.

Although various distance measurement techniques are available as distance criteria, the most commonly used Euclidean distance measurement is. This distance measurement is the preferred distance measurement for measuring proximity in classification and clustering analyzes (Taşcı and Onan, 2016). Euclidean distance measurement:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ip} - x_{jp})^2} \qquad (3.1)$$

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^{p}((x_{i1} - x_{j1})^2} \qquad (3.2)$$

Other distance measures such as Chebyshev, Minkowski, Manhattan, Mahalanobis are also preferred.

Another important parameter is the assignment of the weight value. Neighboring data, which is closer to the data to be classified, is intended to benefit from the majority vote.

The method of determining the weight values used is calculated with formulas $\frac{1}{d}$ or $\frac{1}{d^2}$ to represent the distance between neighbors d(Doad and Bartere, 2013).

4. MODEL PERFORMANCE IMPROVEMENT METHODS

Mean squared error (MSE), root mean squared error (RMSE), mean percentage error (MPE), and R^2 criteria are used for performance assessments of the algorithms used in the application.

The error statistics show how closely the estimates obtained by the algorithm are closer to the actual values. The lowest error statistic value shows the model that best matches data. The R^2 value is between 0-1. as you approach 1, regression indicates that the model alignment is good.

The statistical indicators and the R^2 formula are listed below.

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

$$MPE = \frac{100}{n}\sum_{j}^{n} \frac{e_j}{Aj}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$

5. PRACTICE THEORY

This study uses daily data from the NASDAQ index between 2016-2021. The data was obtained from investing.com and the scliti learn library was used by writing the application in python. The series is tested for stagnation by performing the ADF, PP, and KPSS unit root tests. Once the level is no longer stable, the series is rendered stable and the series is stabilized.

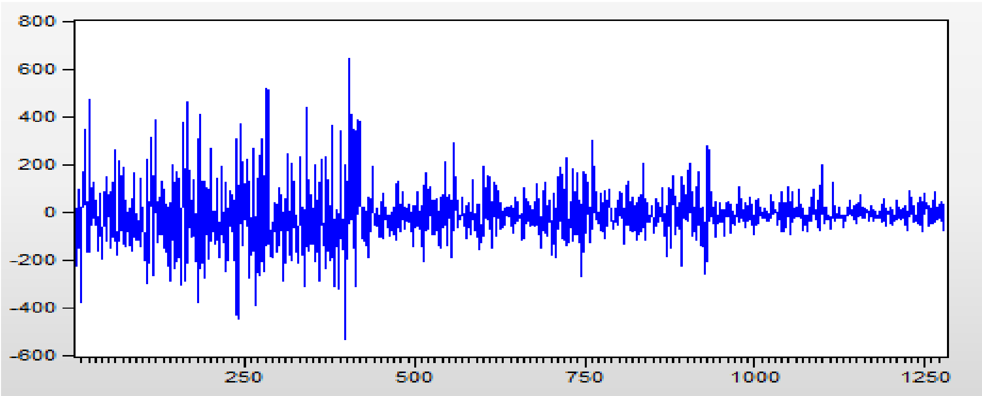
Table 5.1: NASDAQ return series unit root tests

ADF	-3,435***	-2,864**	-2.568*
PP	-3,435***	-2,863**	-2,567*
KPSS	0,739***	0,463**	0,347*

Note: ***, **, * indicates that empty hypothesis is rejected at levels %1, %5, and %10 respectively. It shows that the empty hypothesis for the KPSS test has not been rejected.

The graphic output for the NASDAQ return series is included in Graph 5.1.

Graphic 5.1: NASDAQ return series



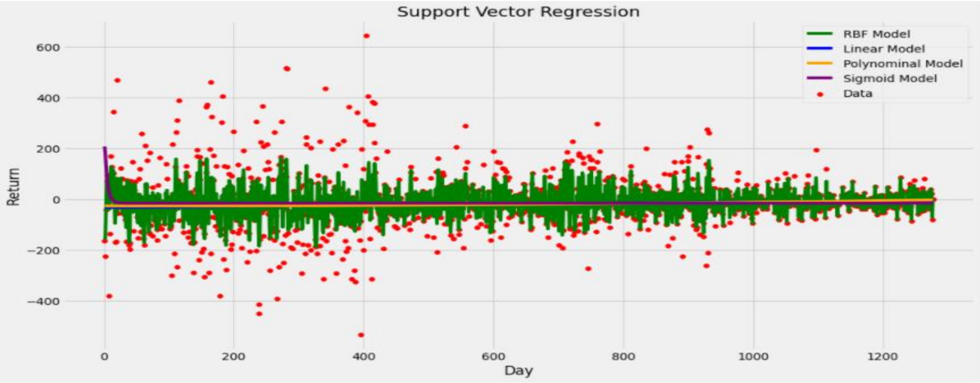
The Jarque-Bera test results applied to the NASDAQ return index are included in Table 5.2.

Table 5.2: Marker statistics

Statistics	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**Jarque-Bera**
Values	-8.047588	109.5707	7.347708	0,654033	1097.74 (0.000)

Data can be processed and predicted by the support vector machine. Graph 5.2 contains a graphic of the series that has been applied to the Support vector machine regression prediction.

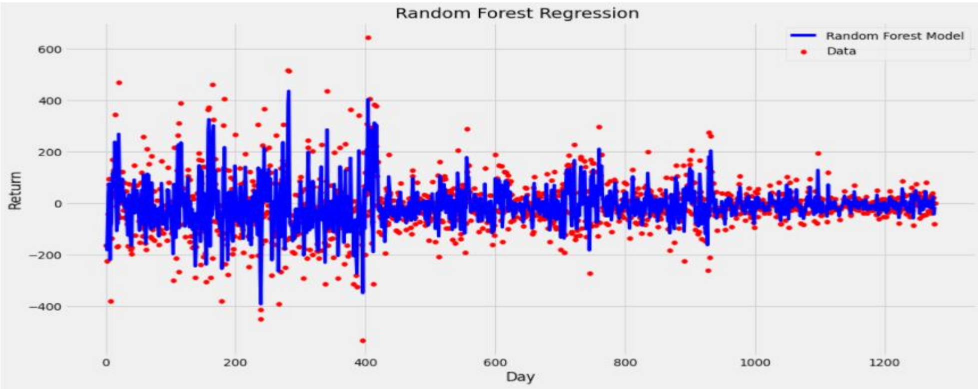
Graphic 5.2: Support vector machine



Graph 2 shows that the best-functioning support vector algorithm is Radial-based. It is observed that estimates are carried out by taking the majority of data for the NASDAQ return series. DVR-Linear, DVR-Polinominal and DVR-Sigmoid models are not working.

By estimating with the Random Forest algorithm, graph 5.3 is obtained.

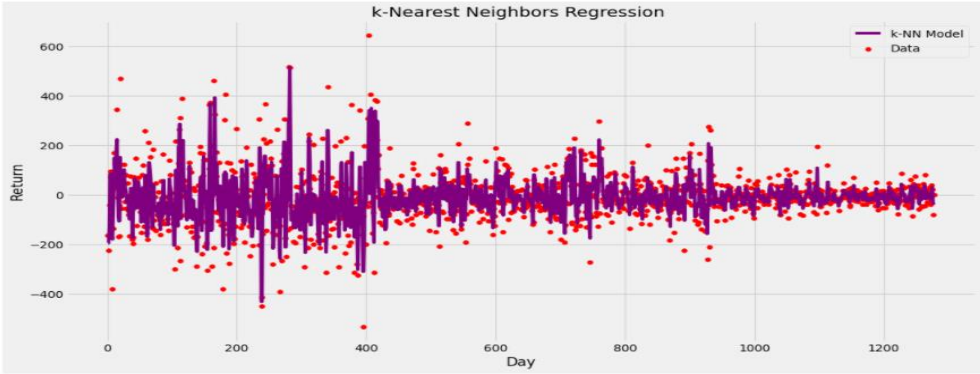
Graph 5.3: Random Forest



Graph 5.3 shows that the random forest algorithm is running.

The graph obtained by performing the estimate of the k-nearest neighbors algorithm is included in Graph 5.4.

Graph 5.4: k - Nearest Neighbors



Graph 5.4 is analyzed to show that it is working efficiently in the K-NN regression, such as RBF-DVR and Random Forest regression.

The actual closing value for the next day is -0.12. to determine the algorithm that makes the most predictive estimate of -0.12, table 5.3 contains estimates for all regression models.

Table 5.3: Statistical values of estimation algorithms

METHOD	Estimation value	MSE	RMSE	MPE	R^2
Linear-DVR	-4.5992	12103.018	110.013	0.9352	0.0080
Polynomial-DVR	-1.5046	12096.366	109.983	0.8670	0.0076
RBF-DVR	**-0.2203**	**10129.418**	**100.645**	**0.8636**	0.6246
Sigmoid-DVR	-14.97	12046.321	109.755	0.8910	0.0040
Random Forest	-14.5085	11989.817	109.498	0.9391	**0.8220**
K-nearest neighbors	-39.295	15034.495	122.615	0.9312	0.5433

When table 5.3 is reviewed, the lowest value for all specified error statistics is RBF-DVR. The R^2 value is found to be the highest Rassal Forest algorithm, while the other highest value is RBF-DVR. The NASDAQ index, when estimated the next day by using historical yield values, shows that the

RBF-DVR algorithm, with its closest value of **-0.12** and the closest value of **-0.2203.**

6. CONCLUSION

In the constantly changing and globalizing world, it is difficult for investors to predict and invest in the future due to the sudden changes in the market. In the time series, artificial intelligence algorithms are preferred for the future prediction, unlike traditional methods.

The study uses the daily closing values of the NASDAQ index in 2016-2021, and because they are not stable at the serial level, returns are obtained to the series using Linear-DVR, RBF-DVR, Polynomial-DVR, Sigmoid-DVR, Random Forest regression and K-nearest neighbors regression algorithms. Error statistics are taken into account for model performance values. The Linear-DVR, Polynomial-DVR and Sigmoid-DVR from algorithms are seen on graphics that do not work. When the graphs are reviewed, it appears that the RBF-DVR, Random Forest regression and K-nearest neighbor regression are active. While R^2 values provide the highest value for Random Forest regression, the lowest error statistics are RBF-DVR. The RBF-DVR is again the most likely to give the next day's return value. Therefore, RBF-DVR is shown to have clearer results than other algorithms. The RBF-DVR algorithm is preferable for the best predictability in the time series, as maximum profits and minimal losses for investment decisions are important to investors.

With the results in mind, effective work can be done with machine learning algorithms instead of traditional econometric methods for the time series. For the financial time series, it is seen that the RBF-SVR, the choice of RF and k-NN regression, will be correct within the scope of the study and will shed light on the future estimated work.

REFERENCES

Alhnaity, B., & Abbod, M. (2020). A New Hybrid Financial Time Series Prediction Model. *Engineering Applications of Artificial Intelligence*, *95*, 103873.

Arslankaya, S., & Toprak, Ş. Makine Öğrenmesi ve Derin Öğrenme Algoritmalarını Kullanarak Hisse Senedi Fiyat Tahmini. *International Journal of Engineering Research and Development*, *13*(1), 178-192.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In *Ensemble Machine Learning* (pp. 157-175). Springer, Boston, MA.

Doad, P. K., & Bartere, M. M. (2013). A Review: Study of Various Clustering Techniques. *International Journal of Engineering Research & Technology*, *2*(11), 3141-3145.

Filiz, E., Karaboğa, H. A., & Akoğul, S. (2017). Bıst-50 Endeksi Değişim Değerlerinin Sınıflandırılmasında Makine Öğrenmesi Yöntemleri Ve Yapay Sinir Ağları Kullanımı. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *26*(1), 231-241.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive Logistic Regression: A Statistical View of Boosting (With Discussion and a Rejoinder by the Authors). *The Annals of Statistics*, *28*(2), 337-407.

Meesad, P., & Rasel, R. I. (2013, May). Predicting Stock Market Price Using Support Vector Regression. In *2013 International Conference on Informatics, Electronics and Vision (ICIEV)* (pp. 1-6). IEEE.

Mitchell, T.M., (1997). Machine Learning.

Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2019). Stock Market Prices Prediction Using Random Forest and Extra *T*ree Tegression. *International Journal of Recent Technology and Engineering.*, *8*(3), 1224-1228.

Taşcı, E., & Onan, A. (2016). K-en Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi. *Akademik Bilişim*, *1*(1), 4-18.

Tay, F. E., & Cao, L. (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega*, *29*(4), 309-317.

Vapnik Vladimir, N. (1995). The Nature of Statistical Learning Theory.

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction Using Machine Learning Techniques. *Procedia Computer Science*, *167*, 599-606.

Zheng, C., Kasprowicz, C. G., & Saunders, C. (2017). Customized Routing Optimization Based on Gradient Boost Regressor Model. *arXiv preprint arXiv:1710.11118*.