Chapter 3

Machine Learning-Based A Comparative Analysis for USA Dollar Index Prediction

Muhammed Fatih Yürük[1]

Abstract

The US Dollar Index is an important indicator of the global economy, as it measures the value of the US dollar against a basket of other currencies. The Dollar Index is used by investors, traders, and decision-makers to inform their investment and trading decisions, as well as to monitor the health of the global economy. In recent years, machine learning techniques have gained popularity in the field of finance for their ability to analyse large amounts of data and provide accurate predictions. This study explores the use of machine learning techniques for predicting the Dollar Index. The study compares the performance of different machine learning algorithms, including Random Forest, Support Vector Machines, and Artificial Neural Networks, in predicting the Dollar Index. The study uses daily data on the Dollar Index from January 2000 to December 2020, which is pre-processed and normalized before being used in the machine learning models. The study finds that machine learning models outperform traditional methods in predicting the Dollar Index. The Random Forest algorithm performs the best among the models tested, with an accuracy of 98.5%. The study also provides a detailed analysis of the feature importance of the input variables in the prediction models, which can help decision-makers understand the factors that affect the Dollar Index. The study concludes that machine learning techniques can provide decision-makers with valuable insights for their investment and trading decisions. The study suggests that future research can explore the use of other machine learning algorithms and input variables to improve the accuracy of the prediction models. The study also highlights the importance of using machine learning techniques in finance and economics, as they can help investors create strong portfolios with little risk.

1 Asist. Prof.Dr., Dicle University, School of Civil Aviation, Department of Civil Aviation Management, mfyuruk@hotmail.com, ORCID: 0000-0001-7429-2278

https://doi.org/10.58830/ozgur.pub105.c739

Introduction

The global economy is heavily dependent on the U.S. dollar, making it the most important currency in the world. Among the reasons why the dollar is such an important currency, the fact that it is the currency used in international trade can be counted as part of the fact that part of the reserves of the central banks of many countries is in dollars. It refers to the dollar reserve amount of central banks and the stability of countries in global financial markets. The strength of the U.S. dollar increases with its use in financial markets. The Dollar Index, also known as DXY, is a measure of the value of the U.S. dollar relative to a basket of foreign currencies. The dollar index refers to the strength of the U.S. dollar against other major currencies. Investors and decision-makers in particular always follow the dollar index, as it can affect international trade, inflation rates, and interest rates. The dollar index informs investors about the global economy. The index includes six major currencies (euro, Japanese yen, British pound, Canadian dollar, Swedish krona, and Swiss franc) and is calculated using the weighted geometric mean of each unit. Considering the exchange rate relations of the six currencies in the created dollar index with the US dollar and the weights of each of them, it will be calculated as in Table 1, given below: Daily dollar index = 50,14348112* EURO/USD-0,576 * USD/JPY0,136 *GBP/USD-0,119 * USD/CAD0,091 * USD/SEK0,042 * USD/CHF0,036 (Yaşar and Yılmaz, 2022).

Table 1. Dollar Index Composition

	Exchange Rate Relationship Coins	**Weight (%)**
1.	EURO/USD	57,6
2.	USD/JPY	13,6
3.	GBP/USD	11,9
4.	USD/CAD	9,1
5.	USD/SEK	4,2
6.	USD/CHF	3,6
	Total	100

Source: (Yaşar and Yılmaz, 2022)

The advance forecast of the dollar index, which is important for investors, traders, country governments, and decision-makers, is desirable. Many techniques and models for forecasting have been developed in the field of finance. These are traditional methods and machine-learning algorithms. Machine learning is a method that is growing in popularity day by day. With machine learning, big data can be trained to obtain prediction outputs. In this period of rapid growth in computer technologies, developing predictive models using machine learning techniques will put decision-makers one step ahead of their competitors. Working with machine learning in finance and economics allows investors to create strong portfolios with little risk. In this study, prediction models have been developed with machine learning techniques, whose importance is increasing day by day around the world.

Artificial neural networks have been developed by scientists who take into account the characteristics of the human brain,creating a mathematical model inspired by the neurophysical structure of the brain. Starting from the idea that to fully model all the behaviour of the brain, its physical components must be modelled correctly, artificial cell and network models have been developed. These developments have revealed a new branch of science. This branch of science is called artificial neural networks, one of the machine learning techniques. Artificial neural networks are considered a different science from the algorithmic calculation methods of today's computers. In this field, to mimic the characteristics of the human brain, a complex network structure is used. Thanks to their learning capabilities, artificial neural networks can be used to analyse, classify, and predict data (Ataseven, 2013). Artificial neural networks are used in a variety of fields. For example, in the medical field, it is used to diagnose and treat diseases. It is also used in many other areas, such as financial forecasting, control of industrial processes, and image recognition. Therefore, artificial neural networks occupy an important place in today's technology. Artificial neural networks (ANNs) are one of the most popular machine learning techniques today. They are widely used in solving various classification and estimation problems. Artificial neural networks are an alternative to classical methods and classical mathematical models, and it is possible to replace these models (Kujawa and Niedbała, 2021). An ANN consists of the union of artificial nerve cells. The cells in this network communicate with each other. The output of one artificial nerve cell is the input of another cell. As shown in Figure 2, cells in the artificial neural network are layered together.

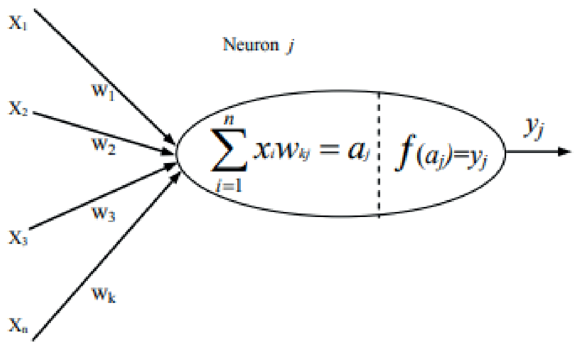

Figure 1. An Artificial Neuron (Yıldız and Yezegel, 2010)

Artificial neural networks are a type of machine learning model inspired by biological nervous systems. The main components of the neural network are neurons, connections, and the learning algorithm. These neural networks are used to process input data and perform specific tasks. The most basic unit of structure in artificial neural networks is called the "neuron." Figure 1 shows the structure of a neuron. A neuron (j) is the basic processing unit of a neural network. All neurons in the network receive a range of inputs (xi) and produce an output (yi). These outputs can either have entered other neurons or gone out of the network (Yıldız and Yezegel, 2010).

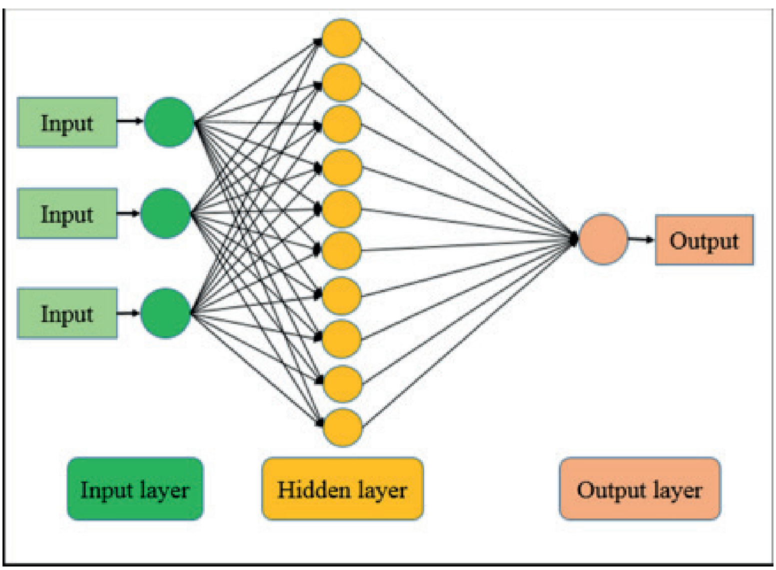

Figure 2. Artificial neural network model diagram (Afan vd., 2015)

As can be seen in Figure 2, the first layer in the ANN where the cells that allow external inputs to be taken into the network are located is the input layer. The layer where the cells take on the task of sending the processed data out of the network is the output layer, and this is the last layer. The layer(s) between these two layers are defined as "hidden layers." When the number of layers is mentioned in an ANN, this refers to hidden layers. The number of input and output layers is not added to this expression. For example, when a 10-layer ANN is specified in Figure 2, it is an artificial neural network with 10 hidden layers and input and output layers located together. Descriptions such as the arrangement of the layers and cells of an ANN and the way they are connected are defined as the architecture of the ANN (Yıldız, 2009). Neurons in artificial neural networks are usually arranged in layers. The input layer receives the input data and contains the neurons that process the input data. Successive layers process input data using various mathematical operations and transfer the results to the next layer, where they transmit them. The output layer produces the final output of the neurons.

Support Vector Machine

One of the machine learning algorithms developed in recent years for solving classification and regression problems is support vector machines. Especially for estimation problems, it produces very successful results compared to traditional techniques. Support vector machines (DVM) have taken their place in the literature as one of the most effective machine learning algorithms applied in the solution of many classification and regression problems with high generalization performance. Support vector machines, which are used in many disciplines, are mostly used in the solution of classification and regression problems in areas such as banking and insurance, medicine, biology, chemistry, social media, industrial sectors, and finance (Ayhan and Erdoğmuş, 2014).

DVM has a solid theoretical foundation within statistical learning theories. DVM has its origins in the Vapnik Chemonvekis Dimension (VC), based on the principle of inherent risk minimization. DVM polygom includes machine learning, a radial-based function network, and two-layer sensor functions (Kaytez, 2012: 68). The support vector machine is capable of dividing data into two or more classes by separation mechanisms in the form of linear in two-dimensional space, planar in three-dimensional space, and hyperplane in multidimensional space (Güran at al., 2014).

Separating space as a high-dimensional linear is important in support vector machines, one of the most important algorithms in machine learning.

With this technique, the data becomes more understandable, so more accurate results can be obtained. Support vector machines (DVM) are supervised learning methods first used by Vapnic (1998) to solve binary classification and regression problems. Aizerman et al. (1964) successfully applied the core idea to broadly margin classes, proving to be powerful tools. Nowadays, DVMs are used in forecasting, financial forecasting, recommendation systems, database marketing, etc. fields (Bordes, 2010: 33).

The nonlinear SVM tries to find a regression function in the hyperspace expressed by $f(x) = w^T\phi(x) + b$. This function is obtained using the "ϵ-insensitive" loss function. The nonlinear SVM can be obtained by solving the following Quadratic Programming Problem (KPP) (Hüseyin and İmamoğlu, 2016):

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} \frac{1}{2}\parallel w \parallel^2 + C(e^T\xi + e^T\xi^*) \\ &\text{s.t. } (\phi(A)w + eb) - Y \leq e\epsilon + \xi,\ \xi \geq 0e \\ &Y - (\phi(A)w + eb) \leq e\epsilon + \xi^*, \xi^* \geq 0e \end{aligned} \tag{1}$$

Here is a predetermined parameter and an arrangement parameter that satisfies the balance between the adaptation of errors and the flatness of the regression function C. ξ and ξ^* are artificial variables indicating whether samples enter the ϵ-tube, e is the unit vector.

Using Lagrange multipliers α and α^*, 1, the dual of KPP (1) is obtained as follows (Hüseyin and İmamoğlu, 2016):

$$\begin{gathered} \max_{\alpha,\alpha^*} -\frac{1}{2}(\alpha^* - \alpha)^T K(A,A^T)(\alpha^* - \alpha) + Y^T(\alpha^* - \alpha) + \epsilon e^T(\alpha^* + \alpha) \\ \text{s. t. } e^T(\alpha^* + \alpha) = 0 \\ 0e \leq \alpha, \alpha^* \leq Ce \end{gathered} \tag{2}$$

After solving KPP (2), we can find $\alpha^{(*)} = (\alpha_1, \alpha_1^*, \alpha_2, \alpha_2^*, \ldots, \alpha_n, \alpha_n^*)$ and the threshold b and then obtain the regression function,

$$f(x) = \sum_{i=1}^{n} (\alpha^* - \alpha)K(x_i, x) + b \tag{3}$$

Here, $K(x_i, x) = (\phi(x_i) \cdot \phi(x))$ represents the kernel function and gives the dot product in hyperspace. α and α^* are Lagrange multipliers and they satisfy $\alpha_i\alpha_i^* = 0, i = 1,2, \ldots, n$.'1. The function f(x) is determined only by the samples (support vectors) with Lagrange multipliers $\alpha_i \neq 0$ or $\alpha_i^* \neq 0$. Furthermore, $A = (x_1, x_2, \ldots, x_n)$ ve $Y = (y_1, y_2, \ldots, y_n)$ denote the inputs and outputs of the training set, respectively (Hüseyin and İmamoğlu, 2016).

KNN (K-Nearest Neighbors) Algorithm

The K-nearest neighbors method (KNN) first came to the fore in the early 1950s. For large training sets of data, the implementation of this algorithm was quite time-consuming. It wasn't until the 1960s that it was widely used. In the 1960s, with the development of computer technology, it began to be widely used (Han et al., 2012). The KNN algorithm is a non-parametric method often used for classification and regression problems (Hu et al., 2016). The KNN algorithm is based on the idea that the outcome of an event will be the same as the outcome of the events closest to it. Through the training set based on past observations entered into the system, dependent variables are determined, which are the result of each element of the data. Prospective predictions will be equal to the average of the results of existing events and the results of the closest elements in the training dataset. Usually, the closest observations are defined as those with the smallest Euclidean distance to the data point under consideration. The Euclidean distance between observations can be found in the example of the 2-dimensional solution depending on the linear distance xi in the x plane and the linear distance yi on the y plane. In regression problems, the dependent variable value to be predicted can be calculated as the arithmetic mean of the dependent values of the optimum k number of neighboring independent values of the independent variable with the estimator (Altunkaynak et al., 2020).

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \qquad (4)$$

Apart from the Euclidean distance, different distance calculation criteria such as Manhattan, Minkowski, and Chebyshev can be used. The functions for the respective distance values are shown in the following equations (Dilki and Başar, 2020):

$$\text{dist}_{\text{minkowski}}\ (x_1 x_2) = \sqrt{\sum_{i=1}^{n}\ |x_{1i} - x_{2i}|^2}$$
$$\text{dist}_{\text{manhattan}}\ (X_1 X_2) = \sum_{i=1}^{n}\ |x_{1i} - x_{2i}|$$
$$\text{dist}_{\text{Chebyshev}}\ (X_1 X_2) = \max_i |x_{1i} - x_{2i}|$$

Literature

It is seen that many studies on machine learning and techniques have been done in the literature. In this section, studies in many branches of science are examined. Takcı (2023) used the KNN algorithm, one of the machine learning algorithms, in medical science. Machine learning algorithms are also widely used in the field of finance. In one of these studies, Arslankaya and Toprak (2021) used polynomial regression and random forest regression from machine learning techniques and recurrent neural networks (RNN) and long-short-term memory (LSTM) from deep learning methods in their stock price forecasting. As a result of the study, it was stated that the best result was the random forest regression model, and the worst result was the polynomial regression model. Filiz et al. (2017) made classifications using the k nearest neighbors algorithm (k-NN), a naive (simple) Bayesian classifier, the C4.5 classification algorithm, and artificial neural networks (ANN) by using the factors affecting the BIST-50 index. As a result of the study, it was stated that the best classification method was the C4.5 classification algorithm, with a rate of 92.71%.

Aksoy (2021) predicted financial statement fraud using machine learning methods such as artificial neural networks (ANN), classification and regression trees (CART), support vector machines (SVM), and logistic regression (LR). It tried to predict whether 88 companies traded on BIST had committed fraud in their financial statements by using machine learning methods one year in advance. As a result of the study, prediction accuracy for ANN (96.15%), CART (96.15%), SVM (80.77%), and LR (80.77) was obtained. If a general evaluation is made, it is seen that the prediction models created in the study were successful.

Papuçcu (2019) stated that stock market index forecasting is difficult and interesting and used machine learning methods in index forecasting. It is also seen in other studies that machine learning algorithms are successful in future financial forecasts. The author has discussed the problem of predicting the direction of the movements of the BIST 100 index. Three different machine learning algorithms—artificial neural networks, support vector machines, and the naive Bayesian classifier algorithm were used, and the performances were compared. At the end of the study, it was seen that all three models could be used to capture stock market index movements, while the artificial neural network algorithm was a better classifier.

Yiğiter et al. (2018) tried to predict the price value of a lease certificate using machine learning techniques. Vakıf Portföy company, which issues sukuk in Turkey, was made on lease certificate prices, and daily price data were modelled using the K-Nearest Neighbors (KNN) algorithm. As a result of the study, the success of the models was measured, and the price predictions made for 1, 3, and 5 days ahead were stated to have given very successful results. Jönsson (2020) used machine learning techniques in his study to predict Swedish GDP growth using the nearest neighbors algorithm. As a result of the study, it was seen that the machine learning algorithm gave good results in the Swedish GDP growth forecast. The author states that it is important to use machine learning techniques to make predictions.

Baybuza (2018) used machine learning techniques to predict Russian inflation. LASSO, Ridge, Elastic Net, Random Forest, and Boosting used machine learning techniques in the study. The study reveals that the Random Forest model and the Boosting model are at least as good at predicting inflation as more traditional models such as Random Walk and autoregression. The author also stated that inflation can be predicted more accurately by using machine learning techniques and algorithms. Carbonneau et al. (2007) compared the performance of machine learning techniques with traditional techniques. A representative set of traditional and machine learning-based prediction techniques were applied to companies' data, and the accuracy of the methods was compared. As a result of the study, it was stated that the performance of machine learning techniques was not better than traditional techniques. However, using a support vector machine (DVM) trained on multiple demand series has produced the most accurate estimates. Gareev (2020) tried to estimate the growth rate of quarterly gross fixed capital formation in Russia using machine learning methods.

Material-Method

This study aims to predict the US dollar index by using some methods and algorithms of machine learning. The study used five variables, as given in table 2, to estimate the U.S. dollar index. Models were created using weekly data between 7/18/2012 and 02/04/2023. 665 weeks of data were entered into the system.

Table 2. Variables Used in the Study

Code	Data Date Range	Continuous targets	Continuous inputs (predictor)	Source
Crude oil	7/18/2012-02/04/2023	US Dollar Index	Crude Oil WTI	https://www.investing.com/
Nasdaq	7/18/2012-02/04/2023	US Dollar Index	NASDAQ Composite (IXIC)	https://www.investing.com/
Diji	7/18/2012-02/04/2023	US Dollar Index	Dow Jones Industrial Average (DJI)	https://www.investing.com/
SP500	7/18/2012-02/04/2023	US Dollar Index	S&P 500 (SPX)	https://www.investing.com/
US-10Y	7/18/2012-02/04/2023	US Dollar Index	United States 10-Year Bond	https://www.investing.com/

Within the scope of the study, three machine-learning techniques were used. Artificial Neural Networks, Support Vector Machines, and the K-Nearest Neighbors Algorithm (KNN) are the techniques used in the study.

Results

Neural Networks Model Results

In machine learning techniques, the model needs to be trained. Therefore, in the artificial neural network model created, 70% of the data is reserved for training and 30% for testing.

Table 3. Artificial Neural Networks Models

Index	Net Name	Training Perf.	Test Perf.	Training Algorithm	Error Funct.	Hidden Act.Funct.	Output Act. Funct.
1	MLP 5-10-1	0.977275	0.969827	BFGS94	SOS	Tanh.	Exponential
2	MLP 5-4-1	0.970723	0.967853	BFGS89	SOS	Tanh.	Exponential
3	MLP 5-10-1	0.978416	0.969811	BFGS 206	SOS	Exponential	Tanh.
4	MLP 5-4-1	0.970112	0.968335	BFGS 80	SOS	Logistic	Logistic
5	**MLP 5-11-1**	**0.978041**	**0.971996**	**BFGS 165**	**SOS**	**Exponential**	**Tanh.**

After the introduction of the data to the system as a training and test set, the best-performing model trials were conducted. As can be seen in table 3, five models were created. As can be seen in the table, the model was created by separating the data into training and test sets. Broydon-Fletcher-Goldfarb-Shanno (BFGS) was used for the training algorithm. The error function Sum of Squares (SOS) was used. The exponential logistic activation function was used in the hidden layers of the ANNs' models. Tanh, logistic, and exponential were used as output activation functions in the models. The network with the highest performance was network number 5. This network is a multilayer sensor (MLP), that is, a model of a feed-forward artificial neural network that produces a series of outputs from a series of inputs. The model consists of 11 hidden layers (5-11-1). When the model was run, the training performance was 0.978041 and the test performance was 0.971996. The training algorithm is BFGS 165, the error function is

SOS, the hidden layer activation function is exponential, and the output activation function is Tanh.

Table 4. Predictions Statistics

Statistics	**5-10-1 (1)**	**5-4-1 (2)**	**5-10-1 (3)**	**5-4-1 (4)**	**5-11-1 (5)**
Minimum prediction (Train)	75.2254	76.6805	74.0959	76.3020	74.0309
Maximum prediction (Train)	113.5585	112.9982	112.3671	110.1118	111.9978
Minimum prediction (Test)	75.2858	76.9515	74.6593	76.5478	74.3406
Maximum prediction (Test)	112.0627	111.4603	111.8874	109.9335	111.6364
Minimum residual (Train)	-6.1048	-6.0133	-5.5688	-6.2121	-5.2889
Maximum residual (Train)	5.1345	6.8291	6.8285	6.2695	5.6767
Minimum residual (Test)	-5.2760	-5.1464	-7.1675	-5.2926	-4.9688
Maximum residual (Test)	7.4591	6.3754	**5.9253**	6.0906	6.1448
Minimum standard residual (Train)	-4.5486	-3.9554	-4.2584	-4.0458	-4.0108
Maximum standard residual (Train)	3.8256	4.4920	5.2216	4.0832	4.3048
Minimum standard residual (Test)	-3.6328	-3.4366	-4.9470	-3.5665	-3.5541
Maximum standard residual (Test)	5.1360	4.2573	**4.0897**	4.1043	4.3952

Table 4 provides the statistical results of the prediction outputs of artificial neural network models. When the results were examined, the maximum residual value of the test results of the 5-10-1 MLP model number 3 was 5.9253 and the maximum standard residual value was 4.0897.

Table 5. Performance Statistics of Models

Performance Statistics	**5-10-1 (1)**	**5-4-1 (2)**	**5-10-1 (3)**	**5-4-1 (4)**	**5-11-1 (5)**
MSE (Mean Squared Error)	4.21	4.48	4.19	4.40	3.90
MAE (Mean Absolute Error)	1.56	1.69	1.55	1.65	1.54
RMSE (Root Mean Squared Error)	2.05	2.11	2.04	2.09	1.97
MAPE (Mean Absolute Percentage Error)	1.74	1.89	1.71	1.84	1.69
R^2 (Determination Coefficient)	0.96	0.96	0.96	0.96	0.97
CORR (Correlation Coefficient)	0.96	0.96	0.96	0.96	0.97

Performance evaluation was performed on the statistical results of the test outputs of five artificial neural network models created in the study (table 5). Statistics used to measure the prediction accuracy (prediction performance) of models (MSE, MAE, RMSE, MAPE, R2, and CORR) were evaluated by calculating. These criteria are frequently used in performance evaluation criteria in the literature. When the performance statistical values of the 5 network models are compared, it is seen that the statistical values of network number 5 (5-11-1) are smaller than the MSE, MAE, RMSE, and MAPE statistical values. When the determination and correlation coefficients of the networks are examined, it is seen that the values of network number 5 are larger. When all these criteria were evaluated, network number 5 showed the best prediction performance. When all criteria have been evaluated (5-11-1), the artificial neural network model will be used in the next part of the study.

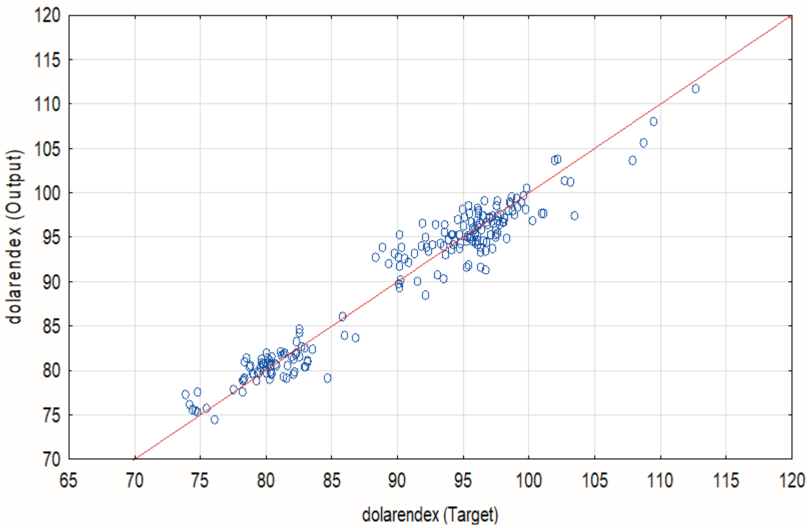

Figure 3. Forecast Graph of Network Number 5 (Test)

In Figure 3, the prediction and input values of network number 5 (test), which shows the best prediction performance from the 5 artificial neural network models created, are given graphically. On the chart, the Y axis contains the values predicted by our model, while the X axis contains the actual values in our model. In other words, the actual data of the dollar index and the dollar index data predicted by the model are positioned on the chart. The blue bubbles appear to be concentrated on the red line. This indicates that the difference between the actual value and the forecast value is small and that the prediction performance of the model is strong.

Support Vector Machine Model Results

At this stage of the study, a prediction model was created with support vector machines, which are machine learning techniques. As in the artificial neural network model, the data of five variables was introduced to the system as given in table 2 to predict the US dollar index. Models were created using weekly data between 7/18/2012 and 02/04/2023. 665 weeks of data were entered into the system. 70% of the data was reserved for education and 30% for testing. The data were randomly distributed by the program. Radial base core (RBF) is used as the DVM core type. A cross-validation method was applied to optimize RBF parameters cost (C) and gamma (γ).

Table 6. Support Vector Machine Results

SVM type	Regression type 1 (capacity=8.000, epsilon=0.100)
Kernel type	Radial Basis Function (gamma=0.200)
Number of support vectors	177 (154 bounded)
Cross-validation error	0.018
Observed mean	90.72519
Predictions mean	90.76877
Observed S.D.	8.77542
Predictions S.D.	8.65718
Mean squared error	5.25691
Error means	-0.04358
Error S.D.	2.29813
Abs. error mean	1.82932
S.D. ratio	0.26188
Correlation	0.96533

Table 6 shows the structure of the support vector machine from the machine learning techniques created for predicting the dollar index. After the test phase of the support vector machine, the core type of the best prediction model was determined as RBF and gamma: 0.200. While the number of support vectors is 177, the result of the cross-validation method used to prevent overfitting is 0.018. In addition, the model has a capacity of 8000 and an epsilon of 0.100. With these values, the support vector machine model with the best prediction performance was created. Other statistical information about the model is also shown in the table.

Table 7. ***DVM Performance Statistics***

Performance Statistics	DVM
MSE(Mean Squared Error)	10.67
MAE (Mean Absolute Error)	2.51
RMSE(Root Mean Squared Error)	3.26
MAPE(Mean Absolute Percentage Error)	2.78
R^2	0.96
CORR	0.96

The statistics used to measure the prediction accuracy (prediction performance) of the model are shown in table 7. MSE, MAE, RMSE, and MAPE criteria and the estimation performance of the support vector machine model were compared. MSE: 10.67, MAE: 2.51, RMSE: 3.26, and MAPE: 2.78 results were obtained. When the model is evaluated alone, statistically successful prediction results are obtained.

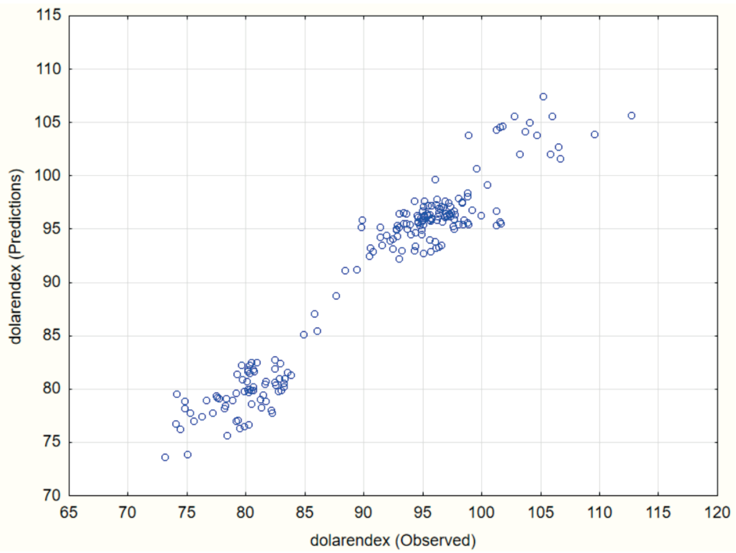

Figure 4. DVM Forecast Chart

After statistically examining the forecast of the dollar index of the support vector machine, the graph of the model was created. On the X-axis are the actual values, while on the Y-axis are the forecast values of the dollar index (figure 4). When the chart is evaluated in general terms, it is seen that the actual and forecast values are close to each other. This indicates that the prediction performance of the model is strong.

Results of the Nearest Neighbors Algorithm Model– (K-Nearest Neighbors – KNN)

In this step of the study, a model was created with the nearest neighbor algorithm, another machine learning algorithm. As in the other steps of the study, the data for five variables was introduced to the system to predict the US dollar index. Models were created using weekly data between 7/18/2012 and 02/04/2023. 665 weeks of data were entered into the system. 70% of the data was reserved for education and 30% for testing. Table 8 gives the statistical information for the model created using the KNN algorithm.

Table 8. Statistical Information of the Model

Mean	91.036
Min	74.105
Max	112.749
Range	38.644
Variance	76.619
Standard Deviation	8.753
Standard Error of Mean	0.340
Median	93.768
Mode	77.026

When the statistical information of the model created by the KNN algorithm is examined, it is seen that the standard deviation is 8.753 and the standard error of the mean is 0.340.

Table 9. Statistical Results of KNN Algorithm Prediction Performance

Performance Statistics	**KNN**
MSE (Mean Squared Error)	0.48
MAE (Mean Absolute Error)	0.50
RMSE (Root Mean Squared Error)	0.69
MAPE (Mean Absolute Percentage Error)	0.55
R^2	0.99
CORR	0.99

The statistical results of the prediction performance of the test set of the KNN algorithm machine learning model is shown in table 9. The evaluation criteria applied in the other steps of the study were applied within the KNN algorithm model. When the performance statistics are examined, it is seen that MSE: 0.48, MAE: 0.50, RMSE: 0.69, and MAPE: 0.55 are the values. In general, when evaluated for the KNN algorithm, the actual values of the dollar index and the prediction values of the model are very close to each other. The KNN algorithm is very good at predicting the values of the input variables in the machine learning model.

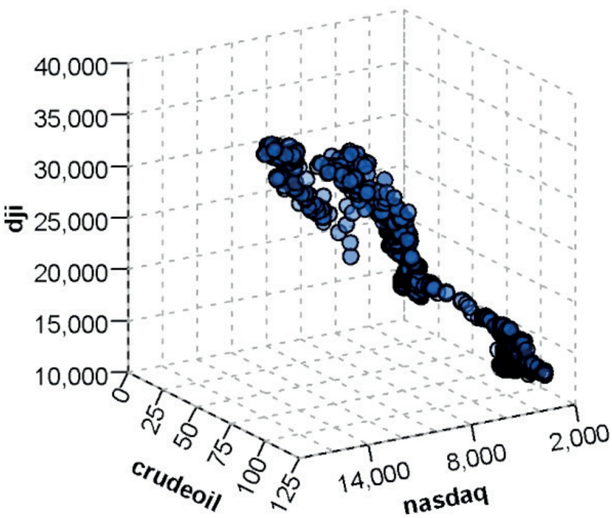

Figure 5. KNN Predictor Space

The prediction space of the KNN algorithm machine learning model is shown in figure 5. The figure is a lower dimensional projection of the predictor space with a total of five predictors. The model is built with three selected predictors and K = 3.

Aggregate Evaluation of the Prediction Performance of Machine Learning Models

Table 10. Statistical Results of the Performance of All Models

Performance Statistics	ANN	DVM	KNN
MSE (Mean Squared Error)	3.90	10.67	**0.48**
MAE (Mean Absolute Error)	1.54	2.51	**0.50**
RMSE (Root Mean Squared Error)	1.97	3.26	**0.69**
MAPE (Mean Absolute Percentage Error)	1.69	2.78	**0.55**
R^2	0.97	0.96	**0.99**
CORR	0.97	0.96	**0.99**

Until this step of the study, the dollar index was estimated using three machine-learning techniques. Table 10 shows the statistical results of the prediction performance of the models. Statistical measurements used to measure prediction accuracy, which is the most commonly used in the literature, were used. Mean squared error, mean absolute error, root means squared error, and mean absolute percentage error criteria were used to evaluate the prediction performance of the three models. When the statistical results were examined, it was found that the prediction performance of all models was very good. Within the three models, the KNN algorithm appears to perform better than machine learning.

Conclusion

The dollar is the most widely used currency in global trade. This currency, which is used at the highest rate in invoicing for imports and exports, further increases the strength of the dollar. The increase in the value of the dollar against other currencies means that the dollar index increases. The change in the dollar index also affects the macro- and microeconomic rates of the countries. The ability to predict the dollar index, which can affect inflation and interest rates, gives the authorities strength in the decision-making process. Forecast models are important for investors to make the right investment decisions.

In the literature, it is seen that financial estimation studies are carried out with traditional and machine-learning algorithms. In this study, predictive models were created using machine learning techniques. In the machine learning studies conducted in the field, it is seen that the prediction performances are high. In this study, it was concluded that the prediction performance was high in artificial neural networks, support vector machines, and k-nearest neighbor algorithm models. If a comparison is made between the three models, it is determined that the model created by the KNN algorithm is more successful. This successful forecasting performance has shown that it will be possible to find instruments that will provide high returns in the field of finance by using machine learning techniques. Country managers, on the other hand, will have the opportunity to take more accurate steps by predicting the change in economic rates with machine learning or artificial intelligence methods. Machine learning methods are contemporary approaches that can be used in all areas of science. By integrating this approach into all units throughout the country, it will be possible for the country to progress economically. In economic terms, progress means the formation of a more stable market and the accompanying increased level of prosperity.

References

Afan, H. A., El-Shafie, A., Yaseen, Z. M., Hameed, M. M., Wan Mohtar, W. H. M., Hussain, A. (2015). ANN-based sediment prediction model utilizing different input scenarios. *Water resources management, 29*, 1231-1245.

Aizerman, M. A., Braverman, E., Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control, 25*, 821-837.

Aksoy, B. (2021). Predicting Financial Statement Frauds Using Machine Learning Methods and Logistic Regression: The Case of Borsa Istanbul. J*ournal of Finance Letters* (115), 27-58.

Altunkaynak, A., Başakin, E. E., Kartal, E. (2020). Air Polution Prediction With Wavelet K-Nearest Neighbour Method *Uludağ University Journal of The Faculty of Engineering (UUJFE) 25*(3), 1547-1556.

Arslankaya, S. & Toprak, Ş. (2021). Using Machine Learning and Deep Learning Algorithms for Stock Price Prediction. *International Journal of Engineering Research and Development, 13*(1), 178-192.

Ataseven, B. (2013). Forecastıng By Usıng Artıfıcıal Neural Networks. *Öneri Journal, 10*(39), 101-115.

Ayhan, S. & Erdoğmuş, Ş. (2014). Kernel Function Selection for the Solution of Classification Problems via Support Vector Machines. *Eskişehir Osmangazi University Journal of Economics and Administrative Sciences 9*(1), 175-201.

Baybuza, I. (2018). Inflation Forecasting Using Machine Learning Methods. *Russian Journal of Money and Finance, 77*(4), 42-59.

Bordes, A. (2010). *New Algorithms for Large-Scale Support Vector Machines.* (PhD Thesis), l'Universitö Paris, Paris.

Carbonneau, R., Vahidov, R., Laframboise, K. (2007). Machine Learning-Based Demand Forecasting in Supply Chains. *International Journal of Intelligent Information Technologies (IJIIT), 3*(4), 40-57.

Dilki, G. & Başar, Ö. D. (2020). Comparison Study Of Distance Measures Using K- Nearest Neighbor Algorithm On Bankruptcy Prediction . *İstanbul Commerce University Journal of Science, 19*(38), 224-233.

Filiz, E., Karaboğa, H. A., Akoğul, S. (2017). Bıst-50 Index Change Values Classıficatıon Usıng Machıne Learnıng Methods And Artıfıcıal Neural Networks. *Çukurova University Journal of Institute of Social Sciences, 26*(1), 231-241.

Gareev, M. (2020). Use of Machine Learning Methods to Forecast Investment in Russia. *Russian Journal of Money and Finance, 79*(1), 35-56.

Güran, A., Uysal, M., Doğrusöz, Ö. (2014). Effects Of Support Vector Machines Parameter Optimization On Sentiment Anaylsis . *Dokuz Eylül Uni-*

versity Faculty of Engineering Journal of Science and Engineering, 16(48), 86-93.

Han, J., Kamber, M., Pei, J. (2012). *Data Mining Concepts and Techniques*: Elsevier Inc.

Hu, L., Huang, M., Ke, S, Tsai, C. (2016). The distance function effect on k-nearest neighbour classification for medical datasets. (Vol. 5 (1), 1304): SpringerPlus.

Hüseyin, İ. & İmamoğlu, S. Z. (2016). Supplier Selection With Support Vector Regression And Twin Support Vector Regression *Dogus University Journal, 17*(2), 241-253.

Jönsson, K. (2020). Machine Learning and Nowcasts of Swedish GDP. *Journal of Business Cycle Research, 16*(2), 123-134.

Kaytez, F. (2012). *Modeling and forecasting of Turkey's long term electricity consumption with least square support vector machines.* (PhD Thesis), Gazi University, Ankara.

Kujawa, S. & Niedbała, G. (2021). Artificial Neural Networks in Agriculture. *Agriculture, 11*(6), 1-6.

Papuçcu, H. (2019). Predicting Stock Market Movement By Using Machine-learning Algorithm . *International Journal of Economics and Administrative Studies* (23), 179-190.

Takcı, H. (2023). Performance-enhanced KNN algorithm-based heart disease prediction with the help of optimum parameters. *Journal of the Faculty of Engineering and Architecture of Gazi University, 38*(1), 451-460.

Vapnic, V. N. (1998). *Statistical learning theory*. United States of America: John Wiley & Sons, Inc.

Yaşar, K. & Yılmaz, E. (2022). Can The Dollar Index Be An Internatıonal Fınancıal Indıcator? Empırıc Revıew On Major Market Indıces. *International Journal of Economics and Administrative Studies* (35), 85-96.

Yiğiter, Ş. Y., Sari, S. S., Karabulut, T., Başakin, E. E. (2018). Estimation of Lease Certificate Price Evaluation Through Machine Learning Method *International Journal of Islamic Economics and Finance Studies, 4*(3), 74-82.

Yıldız, B. (2009). Artificial Intelligence in Financial Analysis (Vol. 1). Ankara.

Yıldız, B. & Yezegel, A. (2010). Fundamental Analysis With Artificial Neural Network. *The International Journal of Business and Finance Research, 4*(1), 149-158.