**Time Series Analysis on Stock Market Data Using Various Algorithms**

**Abstract:**

Predicting stock prices is the most difficult task in the stock market. Due to its qualities and dynamic nature, stock price data is a financial time series data that gets more difficult to predict. [1 Indian stock market prediction using artificial neural networks on tick data Dharmaraja Selvamuthu, Vineet Kumar & Abhishek Mishra]

In the prediction of stock market prices and movements, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are often used.

The algorithms' diverse learning methods provide a unique perspective for a variety of findings.

The ARIMA Model technique is a means to combine technical analysis with financial decision-making.

ANNs are one of the most widely utilized forecasting techniques (Artificial Neural Networks), LSTM (Long Short-Term Memory), SVM (Support Vector Machine), and ARIMA (Auto Regressive Integrated Moving Average) analysis.

Introduction to Forecasting Methods and Models

Forecasting is the system of gathering predictions for the future based totally on historical and present information and the study of trends. The forecasting procedure provides us with a fast and austere way to generate the forecasts for many time series in a single step. Forecasting uses an extrapolative method(s), where the forecasts for a series are only the function of time and past values of the series, not of any other additional variables. A generic example is a review of a few variables of cooking skills at some separate future date. Prediction is a comparable, however extra accepted time period. [ <https://scholar.google.com/citations?view_op=view_citation&hl=it&user=vb9EOUMAAAAJ&citation_for_view=vb9EOUMAAAAJ:HeT0ZceujKMC>]

Stock market is a booming sector of today’s economy; people are investing in stocks for a good return on investment. With the need for more veracity in the trends of values for the stock prices, the trend forecasting becomes more necessary and essentials for stakeholders. [[Sci-Hub | Predicting stock market price using support vector regression. 2013 International Conference on Informatics, Electronics and Vision (ICIEV) | 10.1109/ICIEV.2013.6572570](https://sci-hub.se/10.1109/ICIEV.2013.6572570)]

Machine Learning

Machine studying strategies are divided into categories. The first is supervised learning, in which the training data is a set of labelled instances, each of which is made up of a set of features that is labelled with the right output for that set of features.

For non-stationary data, time-series forecasting is widely employed. Non-stationary data are those whose statistical features, such as the mean and standard deviation, do not remain constant throughout time but instead change. The non-stationary input data used as input to these models is commonly referred to as time-series. Some examples of time-collection consist of the temperature values through the years, inventory fee through the years, fee of a residence through the years etc. So, the enter is a signal (time-collection) this is described through observations taken sequentially in time.

The maximum typically used forecasting strategies consist of ANNs (Artificial Neural Networks), RNNs (Recurrent Neural Networks), LSTM (Long-Short Term Memory), SVM (Support Vector Machine), and ARIMA (Auto Regressive Integrated Moving Average) analysis.

ANN:

Artificial Neural Networks (ANN), commonly simply called Neural Networks (NN), are computer systems that are inspired by the biological neural networks that make up the brains of animals. An ANN is based on a collection of connected units or nodes called artificial neurons that loosely model the neurons in a biological brain. Each connection, like synapses in a biological brain, can carry a signal to other neurons. receives a signal, processes it and can signal neurons connected to it. The "signal" in a connection is a real number, and the output of each neuron is calculated by a nonlinear function of the sum of its inputs. The connections are referred to as edges. Neurons and edges generally have a weight that adapts as learning progresses. The weighting increases or decreases the signal strength of a connection. Neurons can have a threshold so that a signal is only sent when the added signal exceeds that threshold. are added in layers. Different transformation on the provided input are performed by different layers. Signals travel from one layer (the enter layer) to the closing layer (the output layer), probably whilst traversing the layers many times.

Traditional statistical models that include exponential smoothing, moving average, and ARIMA make its prediction linear. Today, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are widely used to predict stock price movements. Artificial neural networks are widely used to solve many problems because of their versatility. ANN can be viewed as a computational or mathematical model that is inspired by the functional or structural properties of biological neural networks. These neural networks are designed to extract patterns from noisy data. A large sample of data known as the training phase then feeds the data into the network that was not included in the training phase, this phase is known as the validation or prediction phase. The only reason for this procedure is to predict new results. Algorithms provide an accuracy of 99.9% using Levenberg-Marquardt, scaled conjugate gradient and Bayesian regularization.

Achieving moderately correct forecasts of a statistic may be a vital however difficult task. ARIMA and

ANN are 2 wide standard and effective prediction models. ARIMA assumes linear information generation’s function, whereas ANN is best suited for nonlinearly generated time series. But, it's virtually not possible to determine the precise nature of a series and a real-world time series most frequently contains each linear still as skew correlation structures. (https://www.sciencedirect.com/science/article/pii/S1877050915006766)

RNN:

RNN is a class of ANN where connections are established based on directed graphs along a temporal sequence. The first layer is normally a feed ahead neural community observed with the aid of using recurrent neural community layer in which a few statistics it had within side the preceding time-step is remembered with the aid of using a reminiscence function. Forward propagation is carried out on this case. It saves statistics required for its future use. If the prediction is wrong, the getting to know price is hired to make small changes. Hence, making it progressively boom closer to making the proper prediction in the course of the backpropagation. Recurrent neural networks can provide better predictions than LSTM (Long-Short Term Memory).

(<https://jfin-swufe.springeropen.com/articles/10.1186/s40854-019-0131-7#Sec1>)

Both finite impulse and infinite impulse recurrent networks can have additional stored states, and the storage can be under direct control by the neural network. The other alternative to this storage is a network or a graph if that incorporates time delays or has feedback loops. Such controlled states are referred to as gated state or gated memory, and are part of long short-term memory networks (LSTMs) and gated recurrent units. This is also called Feedback Neural Network (FNN). [Recurrent neural network - Wikipedia]

Long-Short term Memory (LSTM):

Neural networks utilized in Deep Learning is composed of various layers linked to every different and paintings at the shape and capabilities of the human brain. It learns from big volumes of records and used complicated algorithms to educate a neural net. The recurrent neural community works at the precept of saving the output of a layer and feeding this returned to the enter with a view to expect the output of the layer.

Units are enforced to study very lengthy sequences. This is a extra well-known model of the gated recurrent system. LSTM is extra benign than different deep gaining knowledge of techniques like RNN or conventional feed ahead due to the fact LSTMs address the evanescent gradient issue. Unlike contemporary model for prediction which uses feed forward neural systems, LSTM uses input associations i.e. Not only does the procedure focus on closing day value for stock market data but also all the data points arrangements throughout the day. Which requires a model which incorporates cross-approval which is achieved by training of the model using the pre-partitioned information. The motivation of tuning the trends of stock, is to explicitly amend the calculation so that it can educate to feature data and calibrate itself.

The LSTM module is composed of a cell, a data door, a front door and a door with a view. The cell collects values over arbitrary time intervals, and the three inputs manipulate the development of records inside and out of the cell. Thus, the predominant benefit of the LSTM is each LSTM unit collects statistics for both, an extended or quick period of time (ergo the name) without explicitly using the activation function inside the recurrent components. This lets in LSTMs to take care of the evaporating slope issue – as the value positioned away withinside the reminiscence cell isn't always iteratively adjusted; the inclination does not disappear while it is modelled by the LSTM model. The paper suggests that the algorithm is able to prove the with minimum loss rate of 0.0024 and if the epoch batch rates are increased then training will be more efficient. [https://www.researchgate.net/publication/348390803\_Stock\_Price\_Prediction\_Using\_LSTM].

SVM**:**

Support Vector Machines are efficient supervised learning algorithms applicable for both classification and regression. It uses hyperplane to separate as a classifier.

In classification problems there are a set number of outputs that a feature set can be labeled as, whereas the output can take on continuous values in regression problems. (Predicting Stock Price Direction using Support Vector Machines Saahil Madge) In Saahil’s paper the problem of stock price forecasting as a classification problem.The function set of a stock’s current charge volatility and momentum, together with the index’s current volatility and momentum, are used to predict whether or not the stock’s price m days in the future will be higher (+1) or lower (−1) than the current day’s price. Specifically, we are solving a binary classification problem.

There are no assumptions made in the dataset and all the numeric problems can be dealt with SVM. The linear separability of the data plays a significant role in deciding the degree of tolerance in SVM. The penalty term that is passed as hyperparameter in SVM when it comes to linearly separable and nonlinear solutions is called 'C', which is called the degree of tolerance. The decision limit depends on a small margin and fewer support vectors. Because of this black box method, the tendency towards overfitting and the very strict calculation, it is a useful method that can be carried out even if its high stability is not impaired by the circuit diagrams. [How Does Support Vector Machine (SVM) Algorithm Works In Machine Learning? Analytics Steps]

ARIMA**:**

An ARIMA model could be a category of statistical models for associate analyzing and prognostication statistic data. It expressly caters to a collection of ordinary structures in time series data, and as such provides a straightforward however, powerful methodology for creating skillful time series forecasts.

ARIMA is a term that stands for Autoregressive Integrated Moving Average. It's a more complex version of the simpler Autoregressive Moving Average, with the addition of integration.

This abbreviation is descriptive, capturing the model's major features.

• AR stands for autoregression. The dependent link between Associate in Nursing observation and a few types of lagged observations is used in this model.

• I stand for "integrated." the process of creating a stationary statistic by differencing raw observations (e.g., subtracting an observation from an observation from a previous time step).

• MA stands for Moving Average. The dependency between an observation and a residual error from a moving average model applied to lagged observations is being used in the model.

Each of the compounds is expressly designated as a parameter within the version. ARIMA (p, d, q) is a widely used notation in which the parameters are replaced with integer values to quickly suggest the unique ARIMA version being used. The parameters of the ARIMA model are defined as follows:

• p: The number of lag observations included in the model, also called the lag order.

• d: The number of times that the raw observations are differenced, also called the degree of differencing.

• q: The size of the moving average window, also called the order of moving average.

A linear regression model is created with the necessary number and kind of terms, and the statistics is prepared using a degree of differencing if you want to make it stationary, that is, to remove fashion and seasonal systems that have a negative impact on the regression model. A charge of zero can be used for a parameter, indicating that the parameter is no longer in use. In this way, the ARIMA version can be set up to behave like an ARMA version, or even a simple AR, I, or MA version. When using an ARIMA version for a temporal collection, the underlying system that generated the observations must also be an ARIMA system. This could also be beneficial appear obvious however, it allows to encourage the desire to confirm the model's assumptions within the raw observations and within the residual forecasting errors from the model.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) version of ARIMA is used for seasonal time collection forecasting. ARIMA and its distinctive versions are primarily based totally at the famous Box-Jenkins precept and so those also are extensively referred to as the Box-Jenkins models. [https://scholar.google.com/citations?view\_op=view\_citation&hl=it&user=vb9EOUMAAAAJ&citation\_for\_view=vb9EOUMAAAAJ:HeT0ZceujKMC]

The **Forecasting of demand using ARIMA model paper** used four parameters namely AIC, SBC, variance and maxi-mum likelihood. The data taken is for food company and prediction for its future demand is made. They used SPSS time series module and a fast maximum likelihood estimation algorithm The aforementioned algorithms were used to find values of the four parameters for different models like (1,1,1), (1,0,1) etc. On which they have based their predictions. The ARIMA model (1, 0, 1) is selected because all the coefficients are significantly optimized than any other models taken into consideration. The IBM SPSS Forecasting was used to then obtain the prediction based on the best parameters selected which had the best optimized values for the 4 parameters from the list of all permuted values P, D, Q i.e. (1, 0, 1) was selected.

For predicting, the ARIMA model was used to banking stock market data in this article. The results are obtained using the MINTAB software. In the period 1993 to 2017, 2000 observations were gathered for each variable from associated databases.

Because several ARIMA models can be created for one column of data using different values of p,d, and q, RMSE is chosen as a criterion for finding the fitting ARIMA model. As a result, the fitted ARIMA model has a lower RMSE.

They have concluded on these observations:

1) The values of p, d, and q are between 0 and 2 solely because these values cannot be negative, and they should not be greater than 2 otherwise the parameter estimation will be useless.

2) The RMSE is set between 4.00 and 5.00 depending on the dataset. As a result, after utilizing the program to construct the dataset, ARIMA (1,1,2) was found to be the best with an RMSE of 1.4.

3) In some circumstances, the ARIMA model is not fitted, indicating that the dataset cannot be estimated, and this should be discarded.

Conclusion

In “ARIMA Model in Predicting Banking Stock Market Data” they have made short term forecast on banking stock market data and collected 200 observations, the best model was selected with the criteria of MSE for short term prediction. The forecasting made in “Forecasting of demand using ARIMA model” paper is dependent on four criteria’s namely SBC, AIC, standard error, and maximum likelihood. This helps in predicting values for a longer time period (January 2016 to October 2016 i.e., 10 months). The criteria’s make the algorithm more feasible for forecasting future demands and reliable guidelines.

Only linear predictions are cultivated in ARIMA modeling alone. It requires combination of other forecasting methods like ANN or RNN to support more robust predictability. This strategy can be used and is appropriate for high-tech market scenarios, particularly for banks, because it gives a significant indicator for the future but is inefficient for not so tightly bounded time series data. There are many factors for different types of datasets to be considered which if taken wrongly may result a varied unrelated misleading output/prediction.

This strategy can be used and is appropriate for high-tech market scenarios, particularly for banks, because it provides a substantial indicator for the future. The approach was designed for short-term forecasting and is not suitable for long-term forecasting. Other forecast horizons for stock market data, such as industrial data, can be investigated in the future. Creating new models that combine qualitative and quantitative methodologies to generate accurate forecasts and improve forecast accuracy in the future. Testing it with a neural network technique and compare it to ARIMA's results to see if the ANN's power in the food industry can be confirmed. In addition, creating an ARIMA-radial basis function (RBF) combination can help achieve high accuracy.

References:

<https://scholar.google.com/citations?view_op=view_citation&hl=it&user=vb9EOUMAAAAJ&citation_for_view=vb9EOUMAAAAJ:HeT0ZceujKMC>

Time-Series Forecasting: Predicting Stock Prices Using An ARIMA model  [Serafeim Loukas](https://seralouk.medium.com/?source=post_page-----2e3b3080bd70--------------------------------) [Jul 23, 2020](https://towardsdatascience.com/time-series-forecasting-predicting-stock-prices-using-an-arima-model-2e3b3080bd70?source=post_page-----2e3b3080bd70--------------------------------)

A SURVEY ON ARIMA FORECASTING USING TIME SERIES MODEL Z. Asha Farhath1 , B. Arputhamary 2 , Dr. L. Arockiam3 IJCSMC, Vol. 5, Issue. 8, August 2016, pg.104 – 109

Time-Series Forecasting: Predicting Stock Prices Using An LSTM Model [Serafeim Loukas](https://seralouk.medium.com/?source=post_page-----6223e9644a2f--------------------------------) [Jul 10, 2020](https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f?source=post_page-----6223e9644a2f--------------------------------)

Financial time series forecasting using support vector machines September 2003 [Neurocomputing](https://www.researchgate.net/journal/Neurocomputing-0925-2312) 55(1-2):307-319

Financial time series forecasting using support vector machines Kyoung-jae Kim∗ Department of Information Systems, College of Business Administration, Dongguk University, 3-26, Pil-dong, Chung-gu, Seoul 100715, South Korea Received 28 February 2002; accepted 13 March2003

Predicting Stock Price Direction using Support Vector Machines Saahil Madge Advisor: Professor Swati Bhatt

How Does Support Vector Machine (SVM) Algorithm Works In Machine Learning?: https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning

Selvamuthu, D., Kumar, V. & Mishra, A. Indian stock market prediction using artificial neural networks on tick data. *Financ Innov* **5,**16 (2019). https://doi.org/10.1186/s40854-019-0131-7

Time series forecasting using artificial neural networks methodologies: A systematic review Future Computing and Informatics Journal, Volume 3, Issue 2, 2018, ISSN 2314-7288, <https://doi.org/10.1016/j.fcij.2018.10.003>. (<https://www.sciencedirect.com/science/article/pii/S2314728817300715>)

Ina Khandelwal, Ratnadip Adhikari, Ghanshyam Verma, Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition, Procedia Computer Science, Volume 48, 2015, Pages 173-179, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2015.04.167>. (<https://www.sciencedirect.com/science/article/pii/S1877050915006766>)

Tealab, Ahmed & Hefny, Hesham & Badr, Amr. (2017). Forecasting of nonlinear time series using ANN. Future Computing and Informatics Journal. 2. 10.1016/j.fcij.2017.05.001.

Pramod, & Pm, Mallikarjuna. (2021). Stock Price Prediction Using LSTM. Test Engineering and Management. 83. 5246-5251.

Time-Series Forecasting: Predicting Stock Prices Using An LSTM Model [Serafeim Loukas](https://seralouk.medium.com/?source=post_page-----6223e9644a2f--------------------------------) [Jul 10, 2020](https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f?source=post_page-----6223e9644a2f--------------------------------) https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f

BASE PAPER

ARIMA Model in Predicting Banking Stock Market Data. Mohammad Almasarweh1 & S. AL Wadi2. Modern Applied Science; Vol. 12, No. 11; 2018 ISSN 1913-1844 E-ISSN 1913-1852 Published by Canadian Center of Science and Education

Fattah, Jamal & Ezzine, Latifa & Aman, Zineb & Moussami, Haj & Lachhab, Abdeslam. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management. 10. 184797901880867. 10.1177/1847979018808673.