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

**Abstract: 300 to 400 words**

### Introduction

The stock market has its biggest challenge of predicting the stock prices. The stock price data represents a financial time series data which becomes more difficult to predict due to its characteristics and dynamic nature. [1 Indian stock market prediction using artificial neural networks on tick data Dharmaraja Selvamuthu, Vineet Kumar & Abhishek Mishra]

### Case description

The use of Support Vector Machines (SVM) and Artificial Neural Networks (ANN) is wide in the prediction of stock market prices and its movements. The different methods of learning implemented by the algorithms gives a unique perspective for various insights. ARIMA Model method is a way to integrate technical analysis for making cognizant financial decision.

### Discussion and evaluation

Most ordinary techniques used in the prognosticate a financial time series are Support Vector Machine (SVM), Support Vector Regression (SVR) and Back Propagation Neural Network (BPNN). In the given abstract, we use the full historical data of a company, analyse it with algorithmic models like LSTM, ARIMA, ANN, Bayesian Regularization We use neural networks based on three different learning algorithms, i.e., Levenberg-Marquardt, Scaled Conjugate Gradient and Bayesian Regularization.

### Conclusion

The accuracy on the historical dataset comes out to be……..respectively, significantly less in accordance with the whole historical data.

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

There are two general classes of machine learning techniques. The first is supervised learning, in which the training data is a series of labeled examples, where each example is a collection of features that is labeled with the correct output corresponding to that feature set [J. Brownlee. A tour of machine learning algorithms.].

The most commonly used forecasting methods include 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.

Time-series forecasting is extensively used for non-stationary data. Non-stationary data are called the data whose statistical properties e.g. the mean and standard deviation are not constant over time but instead, these metrics vary over time.

These non-stationary input data (used as input to these models) are usually called time-series. Some examples of time-series include the temperature values over time, stock price over time, price of a house over time etc. So, the input is a signal (time-series) that is defined by observations taken sequentially in time.

**ARIMA:-**

Basically, means that the time series needs to have a constant mean constant variance over time has no seasonality and so this seems like it satisfies most of those conditions except of course it doesn't have a constant mean over time the mean is of course shifting upward it seems in a linear fashion so we can't use this straight ARMA model which is a shame because if we were to somehow eliminate this trend then we probably could use it because the rest of the conditions would be satisfied that's where the ARIMA model comes in so you use the ARIMA model in situations like this where things seem like they're stationary except for a pesky moving average moving mean that comes in here so Arima stands for autoregressive moving average and the I stand for integrated in this context just means that instead of predicting the time series itself you're going to be predicting differences of the times from one timestamp to the previous timestamp of the basic form of ARIMA model is ARIMA 1 1 1 notice that arma had two parameters P and Q the P being the AR order and the to being the MA order ARIMA has three parameters a P of D the Q the P and Q are the same the P applies to the worker of the AR part the Q is the order of the MA part and the D as you might have guessed is the order of the integrated part in this case what we did was a D equals 1 a difference 1 because we just took the first difference

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 an form that stands for Autoregressive Integrated Moving Average. it's a generalization of the less complicated Autoregressive Moving Average and adds the notion of integration. This acronym is descriptive, capturing the key aspects of the model itself.

Briefly, they are:

• AR: Autoregression. A model that uses the dependent relationship between Associate in Nursing observation and a few varieties of lagged observations.

• I: Integrated. the utilization of differencing of raw observations (e.g., subtracting an observation from an observation at the previous time step) so as to create the statistic stationary.

• MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of those additives are explicitly designated withinside the version as a parameter. A widespread notation is used of ARIMA (p, d, q) in which the parameters are substituted with integer values to fast 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 constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e., to remove trend and seasonal structures that negatively affect the regression model. A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model. Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. This may seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.

For seasonal time series forecasting, a variation of ARIMA, viz. the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used. ARIMA model and its different variations are based on the well-known Box-Jenkins principle and so these are also broadly known as the Box-Jenkins models. [https://scholar.google.com/citations?view\_op=view\_citation&hl=it&user=vb9EOUMAAAAJ&citation\_for\_view=vb9EOUMAAAAJ:HeT0ZceujKMC]

RNN:

RNN Designed to keep the output of a layer, Recurrent Neural Network is fed lower back to the enter to assist in predicting the final results of the layer. 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.

**LSTM:-**

This Recurrent Neural Network tutorial will help you understand what is a neural network, what are the popular neural networks, why we need recurrent neural network, what is a recurrent neural network, how does a RNN work, what is vanishing and exploding gradient problem, what is LSTM and you will also see a use case implementation of LSTM (Long short term memory). Neural networks used in Deep Learning consists of different layers connected to each other and work on the structure and functions of the human brain. It learns from huge volumes of data and used complex algorithms to train a neural net. The recurrent neural network works on the principle of saving the output of a layer and feeding this back to the input in order to predict the output of the layer. Now lets deep dive into this video and understand what is RNN and how does it actually work.

**SVM:-**

Support Vector Machines are efficient supervised learning algorithms applicable for both classification and regression. It is a discriminative classifier that is formally defined by a separating hyperplane.

Support Vector Machines are powerful supervised learning algorithms for both classification and regression. It is a discriminative classifier that is formally defined by a separating hyperplane. So given labelled training data, the algorithm outputs an optimal hyperplane that categorizes new examples.

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 feature set of a stock’s recent price volatility and momentum, along with the index’s recent 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. [Predicting Stock Price Direction using Support Vector Machines Saahil Madge]

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 a hyper parameter in SVM while dealing with both linearly separable and nonlinear solutions is denoted as ‘C’ that is called as Degree of tolerance. Large value of C results in the more penalty SVM gets when it makes a misclassification. The decision boundary will be dependent on narrow margin and less support vectors. Because of this Blackbox method, inclination of overfitting, and very rigorous computation makes it an exploitable method to be performed even though its high stability does not get affected by outliners. [[How Does Support Vector Machine (SVM) Algorithm Works In Machine Learning? | Analytics Steps](https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning)]

References:

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