EasyChair Preprint

№ 10219

Analysis and Forecasting of Stock Market Using Time Series Algorithm

Ananthula Jeevan Kumar, Onteddu Gangadhar Reddy, Badveli Sai Kumar Reddy, Balagonda Sandeep Kumar and Arikala Sri Hari

May 21, 2023

ANALYSIS AND FORECASTING OF STOCK MARKET USING TIME SERIES ALGORITHM

ABSTRACT:

Stock market is the way of investing money and making profits by it. It involves a large number of investors, buyers and sellers. India is the 2nd largest country which has more population making profits out from the stock market. As India is in 2nd place, making predictions in stock market will be a helpful one for all. Anyway, prediction in stock market is a little challenging one, since it is time fluctuating and dynamic in nature. The stock market will be changing in irregular intervals of time. So, our model makes prediction in stock market by employing both TSLM (time series linear model) and Support vector machines. By this we develop a model which is used to make the predictions for our future use. Time series play a crucial role in stock market prediction and it can be used to make short term predictions. We also use the ARIMA (Autoregressive integrated moving average) algorithm too. Finally, the forecasted values are converted to the original scale by applying Trend and Seasonality constraints back. In the paper we survey the pros and cons of using these techniques to predict values.

INTRODUCTION :

Stock market is a component of a free market economy. It allows companies to raise money by offering stock shares and corporate bonds and allows investors to participate in the financial achievements of the companies, make profits through capital gains, and earn income through dividends. In stock market prediction, the aim is to predict the future value of the financial stocks of a company. The most used trend in stock market prediction is by using machine learning which makes the predictions in the current values based on its previous value. A correct prediction in stock market leads to huge profits for the sellers. So, to make a correct and accurate predictions machine learning is the best algorithm. The predicted value will be very close to the actual value, thereby increasing the accuracy. The vital part of machine learning is the data set used. The data set must be a perfect one in order to get an accurate predicted value. Choosing of data set is a very important part for developing a perfect model. We need to select a perfect data set as a little change in the data can make a massive change in the outcome. The data set contains 5 variables: open, close, low, high and volume. Volume is the number of shares passed from one owner to another during the particular span of time. Intrinsic value (true value) is the perceived or calculated value of a company, including tangible and intangible factors, using fundamental analysis. It's also frequently called fundamental value. It is used for comparison with the company's market value and finding out whether the company is undervalued on the stock market or not. When calculating it, the investor looks at both the qualitative and quantitative aspects of the business. It is ordinarily calculated by summing the discounted future income generated by the asset to obtain the present value. The methodologies involving in predicting stock market falls in 3 categories. They are fundamental analysis, technical analysis and technological methods. In this paper we are going to use a time series linear dependent algorithm which falls under technical analysis

MOTIVATION:

Many people try to multiply their savings by investing shares/stocks. Share maket is more commonly used by all. To make pofits they need to predict the trend of market exactly and accurately. So making predictions in stock market helps many people to make profit.

OBJECTIVES:

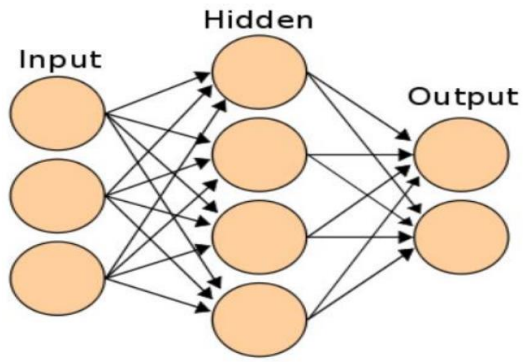
- To understand trend of market accurately.

- To make correct predictions.
- To make it helpful for all.

LITERATURE SURVEY:

1. Artificial Neural Network

It is a technique inspired from human nervous system. It has the ability to predict from large databases. Neural network of multilayer perceptron is used. It consists of an input layer with a set of sensor nodes as input nodes, one or more hidden layers of computation nodes and computation nodes of the output layer. Inputs are passed to input layer simultaneously. The weighted outputs of these units are fed to the next layer of units that make up the hidden layer simultaneously. The weighted outputs of the hidden layers act as an input to another hidden layer. The weighted output of the last the hidden layer acts as inputs to the output layer, which predicts the networks for certain samples. Neural network suffers from Black Box problem (It does not revel the significant weight of every variable). It need to be overtrained since, it has to many of nodes and hidden layers.



2. Hidden Markov Model

It is widely used in stock market sector for the prediction of the data. It makes analysis based on the account close value for a given time line. The probability value gives the stock price trend percentage. In the case of uncertainty, decision-makers make decisions. This model has more accuracy than all but the main problem in it are Evaluation, decoding and learning.

3. Re current Neural Network

Recurrent neural networks use back propagation to learn, but their nodes have a feedback mechanism. Because of this, RNN models can predict a stock price based on recent history and is recurrent. An RNN makes it possible to feed those words in through a much smaller set of input nodes. Which is a major disadvantage for this model.

TIME SERIES LINEAR DEPENDENCY:

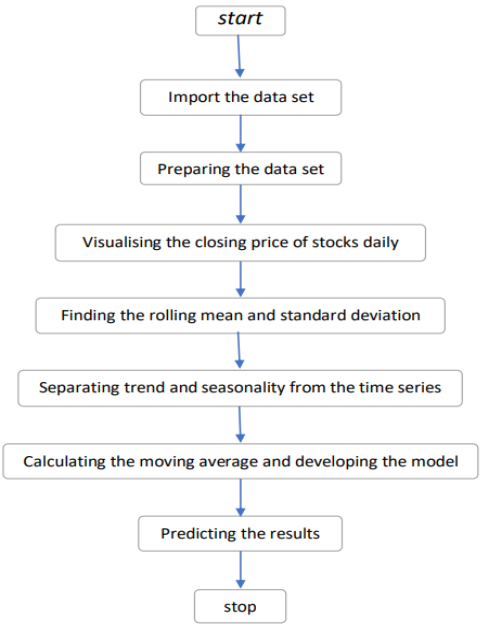
Time series is a type of data that is sampled based on a timebased dimension. We term this data as “dynamic” as we have indexed it based on DateTime attribute. This gives data an implicit order. Static data can still have an attribute that’s a DateTime value but the data will not be sampled or indexed based on the attribute. We apply machine learning algorithms on time-series data and make predictions for the future. These predictions on time-series data are called forecasting. Most procedures for clustering time series look at the similarity of the elements of a set of times series and build a measure of distance between two series by using their univariate features. There are many methods which are useful when we have independent time series and the objective is to cluster them by similarity of their univariate models, in a parametric framework, or by similarity of their periodograms or autocorrelation function. For a set of independent realizations of vectors of stationary time series we measure the disparity between these vectors.

METHODOLOGY:

Time-series forecasting models are the models that are capable to predict future values based on previously observed values. Time-series forecasting is widely 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 that is defined by observations taken sequentially in time.

ALGORITHM OR STEPS:



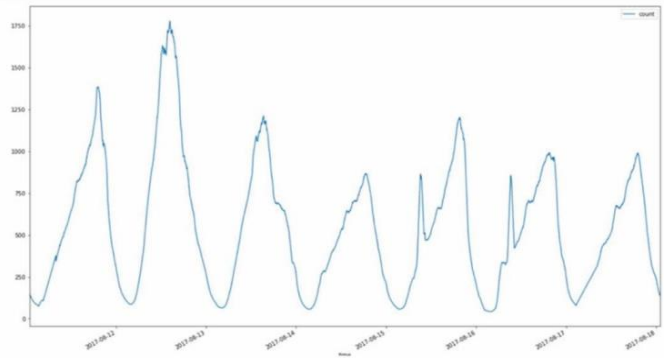
WORKING

Time series forecasting is one of the most applied data science techniques in business, finance, supply chain management, production and inventory planning. Many prediction problems involve a time component and thus require extrapolation of time series data, or time series forecasting. Time series forecasting is also an important area of machine learning (ML) and can be cast as a supervised learning problem. Machine learning methods such as Regression, Neural Networks, Support Vector Machines and Random Forests Import the data set Preparing the data set Visualising the closing price of stocks daily Finding the rolling mean and standard deviation Calculating the moving average and developing the model Predicting the results stop Separating trend and seasonality from the time series can be applied to it. Forecasting involves taking models fit on historical data and using them to predict future observations.

Types of forecasting methods—

Model	USE
Decompositional	Deconstruction of time series
Smooth Based	Removal of anomalies for clear patterns
Moving Average	Tracking a single type of data
Exponential Smoothing	Smooth based model + Exponential Window Function

Moving statistics-



• Frequency – Data is provided at a frequency that is too high to model or is unevenly spaced through time requiring resampling for use in some models.

• Outliers – Extreme outlier values that need to identified and handled.

• Missing – Gaps need to be inter plotted or removed.

Data set-

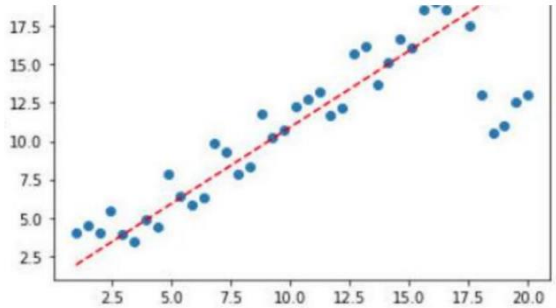
Data set is the input that we give to the model. To make a correct prediction we need to take a perfect or accurate data. Selection of data set plays an important role in developing a model. We need to select with at most care since, our model is totally depended on the data set. Stock Market Turnover Ratio is the data set we used for this model. This information comes from the Federal Reserve Bank of St. Louis. The dataset contains data about the total value of shares traded during certain time periods versus the average market capitalization for that period.

After selecting the data set, the data must be prepared. Data preparation contains three steps.

- Data cleaning
- Outlier reduction
- Data transformation

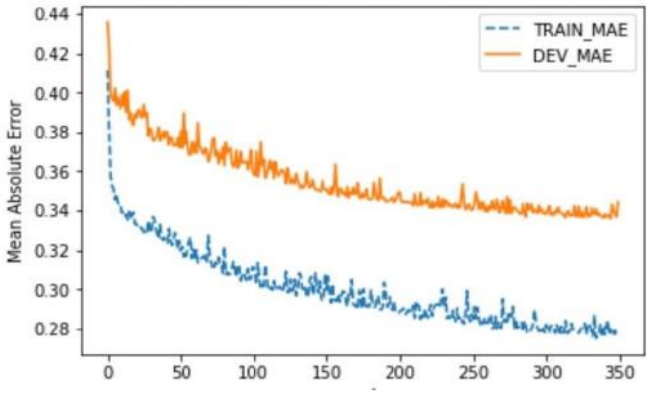
RESULT:

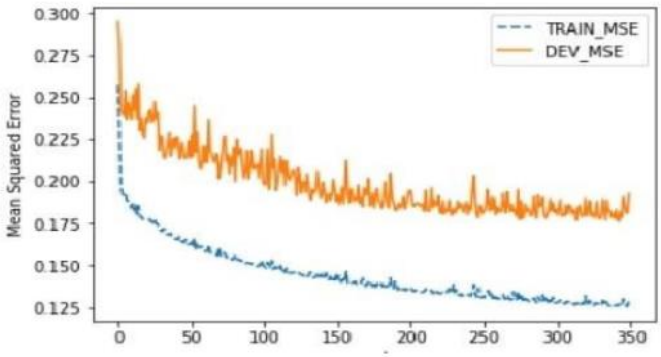
The train data set MSE value and development dataset's MSE started converging. The MAE value also followed the same trend. The model was selected based on both the overfitting metrics and performance metrics. The overfitting metrics was represented with the normalised value of the MSE difference between the train and development data set. And the performance metrics was represented with the normalised value of the MSE difference between train and development dataset. Then the sum of the two normalized value was calculated to find the least value.



Then, the metric values were compared with that of other financial models. However, its performance was nowherenear the time series model's predictive capacity. The time series model MSE value was nearly two thirds of that of other equivalent models.

The MAE metric also showed clear outperformance. The least value of each metric was boldfaced. Here, we can easily notice the how all the metric values of the TSLM model are in boldface. Excluding the 150 assets we have already selected to train our model, we randomly selected 10 assets and generated datasets with identical structures as the ones used in the model training and testing. We then





pass the data into our TSLM hybrid model and evaluate the predictions with the MSE, RMSE and MAE metrics. We iterate this process 10 times to check for model stability. The MSE values of 10 iterations range from 0.1447 to 0.2353. Although there is some variation in the results compared this maybe due to a relatively small sample size, and the outstanding performance ofthe model makes it negligible. Therefore, we may carefully affirm that our TSLM model is robust.

Performance metrics-

	MSE	RMSE	MAE
1	0.1786	0.4226	0.3420
2	0.4597	0.6780	0.5449
3	0.2954	0.5435	0.4423
4	0.4035	0.6352	0.5165
5	0.3079	0.5549	0.4515

MSE - Mean Squared Error

RMSE - Root Mean Square Error

MAE - Mean Absolute Error

CONCLUSION:

Prediction of stock market is not an easy task since, it keeps on changing as time changes. For predicting we need to observe the historical data and need to find the trend of the data. Based on that trend we develop an ideal model which is used to predict the future values. We can't predict the exact value but it can predict the time when we gone have the maximum profit or maximum loss.

REFERENCES:

1) T. J. Urich, E. J. Elton, M. J. Gruber. Are betas best Journal of Finance, 1375–1384, 1978.

2) E. Jondeau, F. Chesnay. Does correlation between stock returns really increase during turbulent periods. Economic Notes, 2001.

3) G. Jenkins, G.E.P. Box. Time series analysis, forecasting and control. Holden-Day, San Francisco, CA, 1970.

4) J. H. Bang, M. Dixon, D. Klabjan. Classification-based financial markets prediction using deep neural networks. Algorithmic Finance,2017.

5) P. Grzegorzewski, P. ladyzynski, K. Zbikowski. Stock trading with random forests, trend detection tests and force index volume indicators. Artificial Intelligence and Soft Computing, Lecture Notes in Computer Science, 2013.

6) G. Swales, Y. Yoon. Predicting stock price performance: A neural network approach. System Sciences, Proceedings of the Twenty-Fourth Annual Hawaii International Conference.

7) F. Zhao. Forecast correlation coefficient matrix of stock returns in portfolio analysis, 2013.

8) W.F. Sharpe. A simplified model for portfolio analysis. Management Science,1963. MSE RMSE MAE 1 0.1786 0.4226 0.3420 2 0.4597 0.6780 0.5449 3 0.2954 0.5435 0.4423 4 0.4035 0.6352 0.5165 5 0.3079 0.5549 0.4515

9) A. Krizhvsky, I. Sutskever, R. Salakhutdinov, N. Srivastave, G. Hinton. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014.

10) T. Tanigawa, K. Kamijo. Stock price pattern recognition - a recurrent neural network approach. 1990 IJCNN International Joint Conference on Neural Networks, 1990.

11) F. Cummins, F.A. Gers, J. Schmidhuber. Learning to forget: Continual prediction with lstm. Technical Report, 1999.

12) M. W. Padberg, E.J. Elton, M. J. Gruber. Simple rules for optimal portfolio selection: The multi group case. Journal of Financial and Quantitative Analysis, 1977.

13) R.A. Oliveira, D.M.Q. Nelson, A.C.M. Pereira. Stock markets price movement prediction with lstm neural networks. In 2017 International Joint Conference on Neural Networks (IJCNN), pages 2017