STOCK PRICE PREDICTION USING TIME-SERIES ANALYSIS

Abhinav
B.Tech Fourth Year
Department of Computer Science
Inderprastha Engineering College
Ghaziabad, Sahibabad India
abhinavch@gmail.com

Anshika Gupta
B.Tech Fourth Year
Department of Computer Science
Inderprastha Engineering College
Ghaziabad, Sahibabad India
anshika.04@outlook.com

Ashutosh Kumar
B.Tech Fourth Year
Department of Computer Science
Inderprastha Engineering College
Ghaziabad, Sahibabad India
ashutoshkumar.15@outlook.com

Shivani Sharma
Assistant Professor
Department of Data Science
Inderprastha Engineering College
Ghaziabad, Sahibabad India
shivani.sharma@ipec.org.in

***Abstract*— In Stock Market Prediction, the point is to foresee the future worth of the monetary loads of an organization. The new pattern in securities exchange forecast advancements is the utilization of AI which makes expectations dependent on the upsides of current securities exchange records via preparing on their past qualities. AI itself utilizes various models to make forecast simpler and credible.**

This project aims to predict the future value of the financial stocks of the companies and aid the beginner to start their journey in Trading in an informed manner, also providing insights about the stocks to the intermediate and expert traders so that they can also make informed decision.

In the project we will be gathering enormous amount of data from the past and try to predict the Future value of the Stock using "Time-series" Analysis, whilst that will be the major guiding parameter for our system, this system will try to gather market emotions through various platforms like Twitter, Journals, News etc. to understand how the stock must be affected and variate the obtained value based on that.

Keywords— Machine Learning Algorithm, Stock Price prediction, time series analysis.

1. INTRODUCTION

As the market is developing, an ever-increasing number of individuals are seeing how their cash ought not be resting, while the entire market is going up with expansion, and putting away their cash to the supplies of organization and developing their portfolio has turned into a pattern and a brilliant decision also.

We see regular new dealers participate in the exchanging scene, and ordinary a large number of them quit as they confronted an immense misfortune in their first speculation. This emerges a default question, WHY? Why new financial investors for the most part go into misfortune while the market is continually developing? The response to this inquiry is likewise exceptionally fundamental, "Un-informed choice" and "Stomach based Decision". Individuals who will more often than not use Vodafone-thought will feel that the item that they use and like ought to likewise develop as it goes, yet that isn't the cases on the lookout and consequently they begin to see their cash going to no end from the exceptionally following day of their excursion in speculation and exchanging.

Stock expectation framework is a framework which will help these amateurs to start their excursion dependent on an educated choice, permitting them more space to learn and develop while the choices are

not only founded on gut, however, is this issue just looked by beginners? NO! experts likewise deal with a similar issue, and they unquestionably can't anticipate the unexpected climbs all by seeing a couple of tweets, and it is outlandish for even Experienced financial backers to perceive how out of nowhere things go to the potential gain or disadvantage. Our framework will assist these prepared investors with witnessing what may and settle on choice dependent on that.

We use the simple procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects.

2. LITERATURE REVIEW

The use of machine learning models in healthcare has increased. Ability of machine learning models to bring out the meaning from data and to prediction is used for early prediction of diseases. Machine learning is used in disease problems to bring out solutions to complex problems. Machine learning is used widely in today's world because of increasing computation power and the availability of large datasets on open-source tools. Quality of transmission (QoT) can give some insights into the model. An attempt has been made to monitor the QoT to determine the physical condition of the model [1]. There is numerous works that has been done related to disease prediction system using different Machine Learning algorithms and achieved different results for different methods in medical field. "Disease Prediction Using Machine Learning over Big Data" has proposed a CNN-MDRP algorithm which combines structured and unstructured data and proved that CNN-MDRP is more accurate than previous prediction algorithm.[2]. Diabetes has been in society for a very long time. Diabetes is further dependent on an individual's body, diet, and way of living [3]. Machine learning models are used in diabetes problems to bring out solutions and to enable early prediction using machine learning models. "Application of Machine Learning Predictive Models in the Chronic Disease "focused on SVM and LR algorithms and evaluates the study models associated with diagnosis of chronic disease. These models are highly applicable in classification and diagnosis of CD [4]. Nonetheless, the presented models lacked details, for instance, Neural Networks parameters such as network size, architecture type, learning rate and back propagation algorithm, etc. In addition, the analysis of the performances is only evaluated in terms of accuracy, which debunks the validity of the presented findings [5].

Moreover, the authors did not take into consideration the bias problem that is faced by the tested algorithms [5],[6]. In illustration, the incorporation of more feature variables could immensely ameliorate the performance metrics of underperformed algorithms [7]. Alghamdi et al. [8] developed an ensemble-based model for predicting incident diabetes. The database used in the research has 32,555 patients. In another study, prediabetes is predicted using machine learning models on the Korean population [9].

3. PROPOSED METHODOLOGY

We utilize a decomposable time series model (Harvey and Peters 1990) with three fundamental model parts: pattern, irregularity, and occasions. They are consolidated in the accompanying condition:

$$Y(t) = g(t) + s(t) + h(t) + Et$$

Here g(t) is the pattern work which models non-intermittent changes in the worth of the time series, s(t) addresses occasional changes (e.g., week by week and yearly irregularity), and h(t) addresses the impacts of occasions which happen on conceivably sporadic timetables more than at least one days. The error term Et addresses any peculiar changes which are not obliged by the model; later we will make the parametric suspicion that Et is ordinarily circulated. This particular is like a summed up added substance model (GAM) (Hastie and Tibshirani 1987), a class of relapse models with conceivably non-direct smoothers applied to the regressors. Here we utilize just time as a regressor yet perhaps a few direct and non-straight elements of time as parts. Demonstrating irregularity as an added substance part is a similar methodology taken by dramatic smoothing (Gardner 1985). Multiplicative irregularity, where the occasional impact is a variable that duplicates g(t), can be cultivated through a log change. The GAM plan enjoys the benefit that it deteriorates effectively and obliges new parts as essential, for example when another wellspring of irregularity is recognized. GAMs additionally t rapidly, either utilizing backfitting or L-BFGS (Byrd et al. 1995) (we favor the last option) so

the client can intelligently change the model boundaries. We are, basically, outlining the guaging issue as a bend fitting activity, which is innately not the same as time series models that unequivocally represent the fleeting reliance structure in the information. While we surrender some significant inferential benefits of utilizing a generative model, for example, an ARIMA, this plan gives various commonsense advantages: {Flexibility: We can without much of a stretch oblige irregularity with numerous periods and let the investigator make various presumptions about patterns.

3.1 Seasonality

Business time series regularly have multi-period irregularity because of the human practices they address. For example, a 5-day work week can deliver results on a period series that recurrent every week, while get-away timetables and school breaks can create outcomes that recurrent every year. To t and gauge these effects we should determine irregularity models that are intermittent elements of t. We depend on Fourier series to give an adaptable model of occasional effects (Harvey & Shephard 1993). Leave P alone the customary period we anticipate that the time series should have (for example P= 365:25 for yearly information or P= 7 for week after week information, when we scale our time variable in days). We can inexact discretionary smooth occasional effects.

3.2 Holidays and Events

Holidays and events give enormous, unsurprising shocks to numerous business timeseries and frequently don't follow an intermittent example, so their effects are not all around demonstrated by a smooth cycle. For example, Thanksgiving in the United States happens on the fourth Thursday in November. The Super Bowl, one of the biggest broadcast occasions in the US, happens on a Sunday in January or February that is hard to announce automatically. Numerous nations all over the planet have significant occasions that follow the lunar schedule. The effect of a specific occasion on the time series is regularly comparative a seemingly endless amount of many years, so it is essential to consolidate it into the estimate.

4. TIME-SERIES ALGORITHMS

Time series is a sequence of main data points in chronological sequence, frequently accumulated in normal spans. Time series analysis can be applied to any factor that changes over the long run and as a rule, typically information focuses that are nearer together are more comparative than those further separated.

Regularly, the parts of time series information will incorporate a trend, seasonality, noise or randomness, a curve, and the level.

Level: When you read about the “level” or the “level index” of time series data, it’s referring to the mean of the series.

Noise: All time-series data will have noise or randomness in the data points that aren’t correlated with any explained trends. Noise is unsystematic and is short term.

Seasonality: If there are regular and predictable fluctuations in the series that are correlated with the calendar – could be quarterly, weekly, or even days of the week, then the series includes a seasonality component. It’s important to note that seasonality is domain specific, for example real estate sales are usually higher in the summer months versus the winter months while regular retail usually peaks during the end of the year. Also, not all time series have a seasonal component, as mentioned for audio or video data.

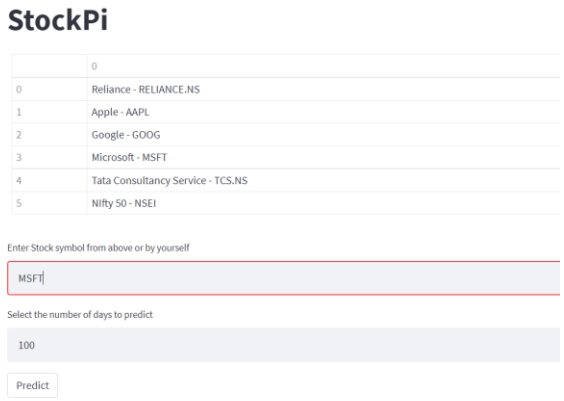
Trend: When referring to the “trend” in time series data, it means that the data has a long-term trajectory which can either be trending in the positive or negative direction. An example of a trend would be a long-term increase in a company’s sales data or network usage.

Cycle: Repeating periods that are not related to the calendar. This includes business cycles such as economic downturns or expansions or salmon run cycles, or even audio files which have cycles, but aren’t related to the calendar in the weekly, monthly, or yearly sense.

5. PROPOSED METHODOLOGY

The suggested solution consists of two key components: 1) a web-based application and 2) a machine learning-based stock prediction system.

The web-application will allow the user to engage with the model and select the stock that they need to see the prediction for, the model itself will collect the raw-data for the time-series analysis apply various filters to it and then come with the given number of days predicted.



7. RESULT

The Stock price prediction uses the time-series analysis with trends and business in the considerations and predicts the future value of the stock closing price in the future. Machine Learning can be a great asset when it comes to solving problems which require speculations although if we need to predict something based on data, Machine learning is the most preferred way of doing so.

Our model is able to create a Simple Moving Average and then able to predict the future values of stock price and give us output so that we can take a informed decision.

VII. CONCLUSION

The major goal of this paper is to predict the values of the stock price of various companies and stock listings by the historical data available. The future scope of this paper are (not limited to): How can we understand the affect of market emotions to the stock price and add it's affect to our analysis and in turn making a model which also checks for the current news, articles, tweets and social media posts and balance the affect in its predictions as well.

We've also built a user-friendly graphical user interface (GUI) to let users engage with the system more effectively. This study demonstrates how a Machine Learning algorithm can be used to easily predict stock price using various factors and models. At the end of the day, we can say that our system has no user threshold because anyone can use it.

8. REFERENCES

[1]. Teo Jašić, Douglas Wood: The profitability of daily stock market indices trades based on neural network predictions: case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965-1999. Applied Financial Economics, 2004, vol. 14, issue 4, 285-297.

[2]. Zhe Liao, Jun Wang: Forecasting model of global stock index by stochastic time effective neural network. Expert Systems with Applications, Volume 37, Issue 1, January 2010, Pages 834-841.

[3]. Eunsuk Chong, Chulwoo Han, Frank C. Park: Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, Volume 83, 15 October 2017, Pages 187-205.

[4]. Taylor SJ, Letham B. 2017. Forecasting at scale.

[5]. Aparna Nayak, M. M. Manohara Pai, Radhika M. Pai, Prediction Models for Indian Stock Market, Procedia Computer Science, Volume 89, 2016, Pages 441-449, ISSN 1877-0509.

[6]. Patil, Rajat. (2021). Time Series Analysis and Stock Price Forecasting using Machine Learning Techniques. 19. 10.1994/Rajat/AI.

[7]. Mehtab, Sidra & Sen, Jaydip. (2020). A Time Series Analysis-Based Stock Price Prediction Framework Using Machine Learning and Deep Learning Models.