2024 3rd IEEE International Conference on Artificial Intelligence for Internet of Things (AIIoT 2024)

Stock Price Prediction using Machine Learning

Marepalli Vishnu Vardhan
School of Electronics Engineering
Vellore Institute of Technology
Vellore, India
vishnuvardhan.mare2021@vitstudent.ac.in

Jaffino G*
School of Electronics Engineering
Vellore Institute of Technology,
Vellore, India
jaffino.g@vit.ac.in

Abstract: **Within the domain of budgetary markets, stock price forecasting has continuously been a challenging but noteworthy undertaking for both pros and financial specialists. The energetic nature of stock markets, which are impacted by a wide range of variables counting financial specialist estimation, geopolitical occasions, and financial figures, makes anticipating more challenging. Through the use of the AutoTS library, a cutting-edge mechanized time arrangement estimating innovation, we aim to improve and speed up the method of stock cost estimation. In contrast to customary machine learning strategies, AutoTS chooses models and adapts to hyperparameters naturally, altering the always moving money related markets. Our investigation, which made use of Yahoo Finance's chronicled stock showcase information, involved meticulous preprocessing of the information to ensure its quality and worldly consistency. The benefits of AutoTS, its adaptability in taking care of complicated designs, and a nitty gritty appraisal of the model's rightness by broad testing on various time periods are secured in detail within the parts that take after. This comprehensive analysis aims to provide insight into the reliability and validity of the AutoTS-based stock cost forecasting system.**

Keywords— Machine Learning, AutoTS, LSTM, Regression, Stock Price Forecasting.

INTRODUCTION:

Driven by the unflinching journey for more exact and reliable disobedience for decision-making within the ever-changing financial markets, the application of machine learning (ML) approaches to the field of stock cost expectation is a charming area of budgetary inquiry. The complexities of always moving advertising conditions, non-linear relationships, and the huge number of exterior components influencing resource values have demonstrated as well much for conventional strategies of stock cost estimation to handle [1][2]. As an effective response to these issues, machine learning shows up, giving the capacity to distinguish complicated designs in huge datasets, alter to changing showcase conditions, and make strides the forecast powers vital for fruitful venture plans. This work points to address the deficiencies of conventional budgetary models and give a more profound understanding of the complexities inalienable in stock cost developments through an investigation of novel calculations and intensive information investigation. This will cultivate headways in prescient precision and choice bolster for speculators and monetary professionals.[3][4]

This demonstration combines chronicled stock cost information, volume exchanged, close-by values, opening and closing costs, and other factors to foresee the long-term cost of the stock. Yahoo (yhfinance), a solid and well-known stage within the monetary industry, given the information used within the research's empirical foundation. Yahoo Fund may be a capable chronicle that gives plenty of authentic and current information on showcase indices, stock prices, exchange volumes, and other money related factors. This stage may be a critical resource that gives us a chance to think about getting to a wide assortment of money related disobedient and ensures a comprehensive examination of past advertising elements. Our investigation is more careful and exact due to Yahoo Finance's user-friendly interface, real-time updates, and extra monetary markers. The approach used to extricate and handle the information is portrayed in profundity within the taking after segments, where we too highlight Yahoo Finance's pivotal role in providing the center dataset for our examination into machine learning-based stock price prediction. [5]

Our study presents a new way to forecast stock costs by using the AutoTS library, withdrawing from the conventional strategies used in prior investigation endeavors. In our journey for moving forward in estimating accuracy, the AutoTS library marks a worldview move indeed in the event that conventional strategies like Random Forest Regression, and Long Short-Term Memory (LSTM) systems have been useful [6]. Modern calculations are used by AutoTS to efficiently ponder and optimize show hyperparameters, making it stand out for its robotized time arrangement determining capacities. The whole determining process is streamlined by this method, which is particularly advantageous because it diminishes the need for human show determination and parameter tuning. We trust to open up the potential of computerized optimization and adaptation by joining the AutoTS library into our investigation system. This might lead to the revelation of idle designs and increase the exactness of stock cost expectations. This novel approach not only it includes to the changing field of money related determining but also emphasizes how pivotal it is to examine state-of-the-art disobedience to make strides in the accuracy and viability of prediction models inside the system of stock advertising elements.

LITERATURE REVIEW:

The body of research on stock price prediction uses a wide range of approaches, with scholars using various models to address the intricacies of financial markets. Previous studies have extensively explored classical linear regression models, harnessing statistical relationships to infer future stock prices. Meanwhile, the application of machine learning techniques, including random forest regression, has gained prominence for its ability to capture non-linear patterns and interactions within market data [7]. Deep learning, particularly LSTM networks, has been a focal point in recent literature, demonstrating proficiency in modelling sequential dependencies and temporal dynamics [8]. Amidst these approaches, our study diverges by incorporating the AutoTS library, a novel automated time series forecasting tool. While traditional methods often necessitate manual selection and tuning of models, AutoTS streamlines this process by automatically identifying optimal configurations, providing a promising avenue for advancing predictive accuracy. This review not only highlights the diverse landscape of stock price prediction but also positions our research within the evolving spectrum of methodologies, emphasizing the need to explore innovative tools for addressing the challenges inherent in forecasting financial markets [9].

METHODOLOGY:

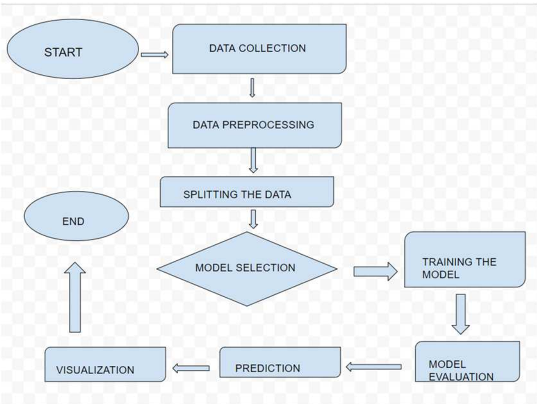

Fig. 1: Methodology

The project follows a systematic methodology for stock price prediction using machine learning. Historical stock data is collected from Yahoo Finance, undergoes preprocessing, and features are engineered to enhance predictive capability. The dataset is split into training and testing sets, preserving temporal order. The AutoTS model selection automates the identification of the optimal time series model, which is then trained, evaluated, and fine-tuned. Performance metrics guide adjustments, and the final model is deployed for real-time predictions. Continuous monitoring, maintenance, and documentation ensure the model's ongoing relevance and accuracy, providing a comprehensive framework for robust stock price forecasting as shown in Fig. 1.

DATA PREPROCESSING:

The information came from the well-known and dependable money-related data site Yahoo Finance, or yhfinance. This site routinely overhauls its information. Information from various companies, such as Google, Amazon, Microsoft, and Apple, is assembled and pre-processed in order to prepare the demonstration. The information is organized into seven columns. They are Date, Open, High, Low, Close, Adj Close, and Volume. This demonstration predicts the closing value of an enterprise for the following ten days. For the most part, the closing cost is the final cost at which a stock exchanges in a standard exchange session.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2012-01-03	1 929333	1 966667	1 843333	1 872000	1 872000	13921500
1	2012-01-04	1 880667	1 911333	1 833333	1 847333	1 847333	9451500
2	2012-01-05	1 850667	1 862000	1 790000	1 808000	1 808000	15082500
3	2012-01-06	1 813333	1 852667	1 760667	1 794000	1 794000	14794500
4	2012-01-09	1 800000	1 832667	1 741333	1 816667	1 816667	13455000

Fig. 2: The first five rows of Google dataset

As shown in Fig. 2, the data was trained from 03-01-2012 to 10-01-2024 and tested from 11-01-2024.

AUTOTS MODEL:

Time series forecasting is facilitated by the Python machine learning framework known as AutoTS, or Automatic Time Series. AutoTS offers access to over 20 model classes, including ARIMA, SARIMAX, Prophet, machine learning, and even deep learning options. AutoTS provides over 30 time-series specific transformations to address seasonality, trends, and other patterns, enhancing data pre-processing for better forecasting. It can be used to anticipate the price of bitcoin as well as stock prices for the ensuing ten days. Among the well-liked attributes of the Python AutoTS library are:

1. Depending on the kind of data you are utilizing, it can be used to determine which time series forecasting model is optimal.
2. Both univariate and multivariate time series are supported by it.
3. It can also manage missing data by filling in and eliminating NaN values, as well as handling outliers.
4. This Python library also offers models that you may utilize for deployment.
5. Uses genetic programming optimization to identify the best time series forecasting model.
6. Provides training and cross-validation for naïve, statistical ML and DL models in every possible configuration of hyperparameters.

```
!pip install autots
from autots import AutoTS
```

Fig. 3: AutoTS Library Setup: Installation and Import Code

The code shown in Fig. 3 guides how to install AutoTS and import AutoTS.

```
model = AutoTS(forecast_length=10, frequency='infer',
               ensemble='simple', drop_data_older_than_periods=200)


model = model.fit(data, date_col='Date',
                  value_col='Close', id_col=None)
```

Fig. 4: Training the data using AutoTS

To train a model using AutoTS, we need to give the forecast length as one of the parameters. If it is set to 10, then it predicts the output for the next 10 days [10]. The frequency is set as "infer" to set the frequency of the Date Time index. Then we use .fit() method to train the data. It takes date_col (Date Column) and value_col (the value model needs to predict) as its inputs, as shown in Fig. 4.

```
prediction = model.predict()
```

```
forecast = prediction.forecast
print("Stock Price Prediction")
print(forecast)
```

Fig. 5: Prediction using AutoTS

After training the model, to predict the output from the model, we use the code shown in Fig. 5 [11].

AUTOTS vs LSTM:

LSTM systems and AutoTS are two unmistakable time arrangement determining strategies, each having points of interest in its own. The computerized time arrangement estimating library AutoTS is essential for its capacity to quicken the show choice and optimization handle. It viably decides the ideal setup based on execution measurements by computerizing the examination of numerous estimating models, enveloping both machine learning calculations and conventional measurable strategies. This concept is requested by analysts as a successful, data-driven strategy since it is mechanized, which radically reduces the need for human intervention.

In successive information, repetitive neural systems (RNNs) of the LSTM sort are especially great at recognizing complex worldly relationships and designs [12][13]. A time arrangement estimating assignment where long-term conditions are pivotal can benefit significantly from the long short-term memory and long-term utilization of LSTM systems [14]. Due to their inborn capacity to capture complex connections and reflect non-linear intelligence, LSTMs are broadly used in various businesses, counting funds [15].

To sum up, AutoTS makes show choice mechanized and adaptable, making it simple and fast to explore with distinctive estimating techniques. On the other hand, LSTM is optimized for consecutive information and is especially great at recognizing non-linear designs and long-term conditions [16].

In a few angles, AutoTS outperforms Long Short-Term Memory (LSTM) systems in time arrangement estimating. Its robotization and adequacy in choosing models are one of its primary points of interest. Without the need for human interaction, AutoTS reliably explores and surveys a wide assortment of estimating models, counting routine factual procedures, and performing machine learning calculations. Since it cuts down on the time and exertion required for trial and fine-tuning, this computerized approach is particularly advantageous when working with an expansive pool of candidate models. Besides, AutoTS is more user-friendly and available for analysts without an incredible bargain of encounter with neural arrangement plans since it alters the inborn complexity of the information without expressly characterizing the demonstration. In contrast to the more complex and parameter-sensitive LSTM systems, AutoTS offers a straightforward and data-driven approach to time arrangement determination, making it an engaging choice, especially when working with distinctive datasets and changing worldly designs.

AUTOTS VS REGRESSION:

Regression models along with AutoTS are the two methods standing up for predicting time series. All of them have benefits in their individual ways. Methodically and data-drivenly doing model configuration and selection using the automated time series forecasting tool AutoTS, can be performed. Overtime, it spans a broad range of forecasting models, adjusting to the intrinsic complexity of the data and the automation process of finely tuning hyperparameters. For this reason, AutoTS is excellently helpful when deciding the optimal forecasting model is quite challenging and the researcher needs a speedy manner to navigate a huge model space.

Contrastingly, time series analysis often finds a more traditional and understandable structure in regression models. Linear regression serves as a prime example in this context. These models thrive on the assumption of a linear connection between input data and the target variable, making them particularly adept when the expected patterns are anticipated to be clear-cut and easily discernible [17][18]. Regression models become the preferred choice when interpretability is a priority, offering insights into the precise influence of each input feature on the expected outcome [19][20].

The specific data type, the intricacy of temporal patterns, and the demands of the given forecasting task all play a role in the decision between AutoTS and regression models [21]. AutoTS excels in automation and flexibility when dealing with complicated and dynamic patterns, while regression models may be a better option when usability, readability, and a clear understanding of feature contributions are seen as important considerations.

RESULTS:

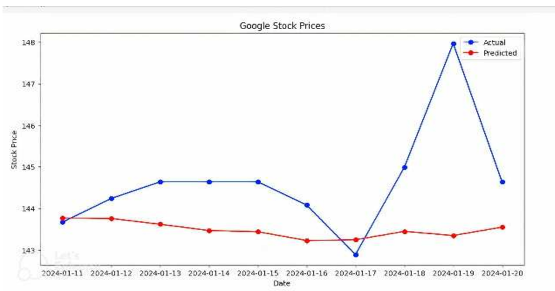

Fig. 6: Google Stock Price Prediction using AutoTS

From Fig. 6, we successfully obtained precise forecasts for Google's stock prices for the following ten days by using the AutoTS model. With the use of automated time series forecasting, the program produced forecasts by identifying trends in past stock data [22][23]. Due to the useful information these forecasts offer regarding the expected direction of Google's stock prices in the near future.

Fig. 7: Amazon Stock Price Prediction using AutoTS

Utilizing the AutoTS model, we have achieved precise predictions for the stock prices of Amazon over the upcoming 10 days as shown in the Fig. 7. Through the application of automated time series forecasting, the model thoroughly examined historical stock data, discerned underlying patterns, and produced reliable forecasts. These prognostications serve as valuable tools for guiding short-term investment decisions, empowering stakeholders to make well-informed choices rooted in the expected trajectory of Amazon's stock prices in the immediate future.

Fig. 8: Intel Stock Price Prediction using AutoTS

Applying the AutoTS model, we have successfully generated accurate predictions for the stock prices of Intel over the next 10 days as shown in Fig. 8. Employing automated time series forecasting capabilities, the model meticulously analyzed historical stock data, identified underlying patterns, and generated reliable forecasts. These predictions serve as crucial insights for making well-informed decisions in short-term investments, providing stakeholders with valuable information on the anticipated trajectory of Intel's stock prices in the immediate future.

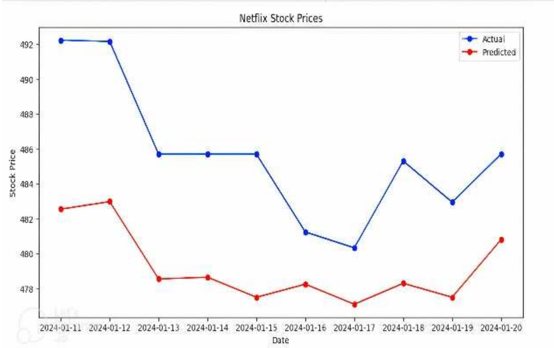

Fig. 9: Netflix Stock Price Prediction using AutoTS

As shown in Fig. 9, we have produced precise forecasts for Netflix's stock prices for the next ten days by using the AutoTS algorithm. With the use of automated time series forecasting, the model produced accurate projections by carefully examining historical stock data and identifying underlying patterns. These forecasts give stakeholders important information about the expected direction of Netflix's stock prices in the future.

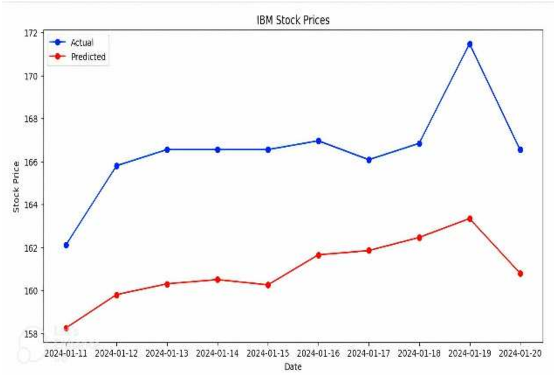

Fig. 10: IBM Stock Price Prediction using AutoTS

Applying the AutoTS model, we have successfully generated accurate predictions for the stock prices of IBM over the next 10 days. Employing automated time series forecasting capabilities, the model meticulously analyzed historical stock data, identified underlying patterns, and produced reliable forecasts as shown in Fig. 10. These predictions serve as crucial insights for making well-informed decisions in short-term investments, providing stakeholders with valuable information on the anticipated trajectory of IBM's stock prices in the immediate future. The following Table I contains the values of MSE, RMSE, and MAE for different companies. These values can be used as performance metrics for the evaluation of the model. Having the lowest RMSE, MSE and MAE values implies that the model has good accuracy .

Table I: AutoTS Model Performance Metrics Table for Different Companies

COMPANY	MSE	RMSE	MAE
Google	2.99	1.729	3.458
Amazon	6.21	2.492	4.984
Intel	0.32	0.565	1.13
Netflix	46.84	6.844	13.688
IBM	33.09	5.75	11.50

CONCLUSION:

In conclusion, this project employed the AutoTS library to construct prediction models for the stock values of various companies. The following performance metrics were discovered: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE). AutoTS's automatic and adaptive characteristics proved helpful in controlling the complicated and dynamic changes in stock values, eliminating any necessity for human model selection or hyperparameter adjustment. Lower values imply better performance, and the final MSE, RMSE, and MAE values offer quantitative insights into the prediction accuracy of the models. The continual implementation of AutoTS across many entities demonstrates its versatility and efficacy in the field of stock price prediction, even if the exact values may change amongst different companies and datasets. These results underline how automated time series forecasting technologies, like AutoTS, can boost the precision and effectiveness of financial forecasts whilst providing investors and practitioners with enlightening advice on navigating the complexities of the stock market.

References:

[1] Adusumilli, R. (2022, July 12). *Machine Learning to Predict Stock Prices - Towards Data Science*. Medium. https://towardsdatascience.com/predicting-stock-prices-using-a-keras-lstm-model-4225457f0233

[2] Schmidt, J., Marques, M. R. G., Botti, S., & Marques, M. A. L. (2019, August 8). *Recent advances and applications of machine learning in solid-state materials science*. Npj Computational Materials. https://doi.org/10.1038/s41524-019-0221-0

[3] Jaffino.G(2021).Grey wolf optimization with deep recurrent neural network. https://ieeexplore.ieee.org/abstract/document/9392541/

[4] Ramakrishnan, M. (2023, July 21). *5 Benefits of Using Machine Learning in Finance*. Emeritus Online Courses. https://emeritus.org/blog/5-benefits-of-using-machine-learning-in-finance/

[5] Yahoo Finance - Business Finance Stock Market News, https://finance.yahoo.com/

[6] Srivatsavaya, P. (2023, October 5). *LSTM — Implementation, Advantages and Diadvantages - Prudhviraju Srivatsavaya - Medium*. Medium. https://medium.com/@prudhviraju.srivatsavaya/lstm-implementation-advantages-and-diadvantages-914a96fa0acb

[7] Chauhan, P. (2023, July 8). *Stock Prediction and Forecasting Using LSTM(Long-Short-Term-Memory)*. Medium. https://medium.com/@prajjwalchauhan94017/stock-prediction-and-forecasting-using-lstm-long-short-term-memory-9ff56625de7

[8] Rodríguez, F. (2021, December 14). *ML & DL—Machine Learning and Deep Learning 101 | Medium*. Medium. https://mafda.medium.com/ml-dl-machine-learning-and-deep-learning-101-2686d93d70d

[9] Sable, A. (2021, April 9). *Time Series Forecasting with Regression and LSTM | Paperspace Blog*. Paperspace Blog. https://blog.paperspace.com/time-series-forecasting-regression-and-lstm/

[10] Jaffino. G (2021). Adaptive rag-bull rider: A modified self-adaptive optimization algorithm for epileptic seizure detection with deep stacked autoencoder using electroencephalogram. https://www.sciencedirect.com/science/article/abs/pii/S1746809420304390

[11] Seshadri R, Gupta N. 2020. auto-ts. pypi.org. https://pypi.org/project/auto-ts/

[12] Kang, E. (2018, June 18). *Long Short-Term Memory (LSTM): Concept - Eugine Kang - Medium*. Medium. https://medium.com/@kangeugine/long-short-term-memory-lstm-concept-cb3283934359

[13] Uddin, S., Khan, A., Hossain, E., & Moni, M. A. (2019, December 1). *Comparing different supervised machine learning algorithms for disease prediction*. BMC Medical Informatics and Decision Making. https://doi.org/10.1186/s12911-019-1004-8

[14] Saxena, A. (2024, February 14). *Introduction to Long Short-Term Memory (LSTM) - Analytics Vidhya - Medium*. Medium. https://medium.com/analytics-vidhya/introduction-to-long-short-term-memory-lstm-a8052cd0d4cd

[15] Hull, G. (2022, December 19). *The Long Short-Term Memory (LSTM) Network from Scratch | Medium*. Medium. https://medium.com/mlearning-ai/building-a-neural-network-zoo-from-scratch-the-long-short-term-memory-network-1cec5cf31b7

[16] R. R. (2023, March 20). *LSTMs for regression - Rupak (Bob) Roy - II - Medium*. Medium. https://bobrupakroy.medium.com/lstms-for-regression-cc9b66677697f

[17] Singh, H. (2021, December 8). *Understanding Random Forests - Harshdeep Singh - Medium*. Medium. https://medium.com/@harshdeepsingh_35448/understanding-random-forests-aa0ccecdbbbb

[18] Sahai, N. (2023, September 6). *Mastering Random Forest Regression: A Comprehensive Guide*. Blogs & Updates on Data Science, Business Analytics, AI Machine Learning. https://www.analytixlabs.co.in/blog/random-forest-regression/

[19] Bhat, S. (2023, January 26). *A Comprehensive Guide to Random Forest Regression - SHRINATH BHAT - Medium*. Medium. https://medium.com/@bhatshrinath41/a-comprehensive-guide-to-random-forest-regression-43da559342bf

[20] Mwiti, D. (2023, September 1). *Random Forest Regression: When Does It Fail and Why?* neptune.ai. https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why

[21] Bajaj, A. (2023, December 8). *Performance Metrics in Machine Learning [Complete Guide]*. neptune.ai. https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide

[22] Jaffino. G (2022). Weighted 1D-local binary pattern features and Taylor-Henry gas solubility optimization based Deep Maxout network for discovering epileptic seizure using EEG. https://www.sciencedirect.com/science/article/abs/pii/S1051200421003882

[23] Nissa, N. K. (2021, December 16). *Stock Price Prediction using Auto-ARIMA - Nuzulul Khairu Nissa - Medium*. Medium. https://nzlul.medium.com/stock-price-prediction-using-auto-arima-5569fcceae59