Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Machine Learning Based Stock Market Trend Prediction and Analysis

Yilin Wang*

Suzhou No.10 High School of Jiangsu Province, Suzhou, China

* Corresponding Author Email: a1anw@stu.hebmu.edu.cn

Abstract. Stock price prediction can help investors to create initial pre-scenarios. The topic of this research is to predict the stock market scenario through machine learning methods. By successfully predicting the stock market situation, the movement of different stocks, etc., the possibility of buying the wrong stocks can be greatly reduced, making it possible to make huge profits by buying and selling stocks. The purpose of the research is threefold. This paper uses a market capitalization weighted index consisting of the most important 40 stocks out of the top 100 stocks with the largest market capitalization on the Paris Stock Exchange as a dataset to analyze the movement of stocks and the possibility of buying them. The stocks are analyzed and subsequently forecasted through data visualization techniques such as candlestick charts and moving average charts. Long Short-Term Memory (LSTM) was used as a benchmark. This study has three main objectives: first, to study the temporal evolution of the closing prices of these stocks; second, to make stock forecasts using candlestick charts, which is the traditional method of stock analysis; and third, to analyze the closing prices of the stocks and the moving average charts, which are studied and evaluated. Stock prediction can reduce risk, guide investment decisions, realize risk management, make market references, assess the value of stocks and enhance strategic vision.

Keywords: Stock Price Prediction; Machine Learning; LSTM; Candlestick Charts.

1. Introduction

Stocks are financial instruments whose origins can be traced back centuries. The earliest stocks date back to the European medieval period when merchants began raising funds for business activities by issuing shares to investors. These shares also represented ownership and equity in the enterprises. Some effective influencing factors in stock price changes are also reasons that cause stock price changes and are difficult to predict [1]. These fluctuations and unpredictability make investors worried. The volatility of stocks can bring both profits and potential negative consequences. Hence, stock prediction has become particularly important. Stock price prediction, as a classic problem, is frequently used for forecasting and research. According to the Efficient Market Hypothesis, predicting stock prices is considered unlikely due to the completely random behavior of stocks. However, some believe that information about stocks can be used to identify trends in price changes through recent or historical prices, analyze underlying reasons, and forecast stock prices [2]. The future performance of the stock market determines the price of stocks [3].

Boosted Regression Trees (BRT) is a statistical model that combines regression trees and boosting methods. It divides the entire dataset into two parts, using 80% randomly selected data for modeling and the remaining 20% for testing model performance. After establishing the BRT model, comparison with observed values and calculation of the Area Under the Curve (AUC) is performed to evaluate the model's quality. A value closer to 1 indicates a better model. In stock prediction, the BRT model has been utilized to forecast stock returns. Through machine learning, researchers have identified a connection between information sets and predictive values, leading to a nonlinear and non-monotonic distribution of investment portfolios in risky assets for investors, providing investment guidance to a great extent [4]. In the realm of machine learning for prediction problems, boosting and bagging are commonly used and popular algorithms. They have been employed in stock prediction, with advancements made in tree-based models leading to gradient boosting and the XGBoost algorithm.

These algorithms are widely used in competitions and can be applied to stock prediction. Currently, deep learning in machine learning has become the mainstream approach for prediction, showing remarkable capability in extracting relevant information from financial time series data [5]. Hidden Markov Models (HMM) generate label sequences, observe data, and visualize dependencies among random variables using Bayesian networks. Overcoming the challenges of using HMM involves identifying hidden parameters from observable parameters and analyzing these parameters, enabling widespread application in various fields such as speech recognition, gesture recognition, and bioinformatics. In stock prediction, some have employed HMM to predict the next day's stock values by analyzing historical data. Researchers use changes in stock value scores and daily high-low values to train continuous HMM and make maximum a posteriori decision. They have applied their methods to multiple stocks and compared their experimental results with existing methods like artificial neural networks in terms of performance [6].

In the realm of stock market prediction, alternative methods have been introduced for forecasting stock market crises. Forecasting crises in the stock market indirectly reflects the trends and fluctuations that stocks are likely to undergo, constituting another facet of stock market prognostication. By leveraging stock index derivatives, researchers have extensively analyzed China's stock market to forecast stock market crises. Their comprehensive study revealed that combining futures and options variables yields optimal model performance and provides strategies for managing crises [7]. Stock price prediction faces an inherent challenge where price differences can fluctuate significantly over a short period. For instance, sudden increases in price volatility may reduce the accuracy and success rate of predictions, leading to forecasting failures. Researchers have addressed this issue by integrating grey theory and fuzzy techniques into Fuzzy Grey Prediction (FGP), aiming to forecast stock prices for the next time period, such as the upcoming hour, and track their trends. This approach offers a novel perspective for enhancing this stock price predictions [8]. Due to the unpredictable and volatile nature of stocks, machine learning algorithms like Support Vector Machine (SVM) are commonly used to address prediction challenges. However, using k-NN (K Nearest Neighbors) regression for market trend prediction is expected to be more accurate. As a result, many researchers opt for the k-NN regression algorithm, finding it highly beneficial across various industries [9]. Using Long Short-Term Memory (LSTM) can make the predictions more accurate, greatly improving the hit rate of stock forecasting [10]. Therefore, this study utilized LSTM for stock prediction and analysis.

2. Organization of the Text

2.1. Dataset Description and Preprocessing

In this study, this study used the dataset from Kaggle [11]. It is called CAC40. It formerly known as the Paris Stock Exchange, is the benchmark stock index of the French stock market. The index represents a capitalization-weighted measure of the 40 most significant stocks among the 100 largest market caps on Euronext Paris. Its acronym stands for Continuous Assisted Trading (Cotation Assistée en Continu) and is used as a benchmark index for funds investing in the French stock market. This research utilized this dataset for analysis and predicted the performance of Atos stock to achieve the experimental research objectives. Obtained from Kaggle, the CAC40 dataset includes the daily opening price, closing price, daily high, daily low, and trading volume for each of the top 40 ranked stocks from 2014 to 2023. This dataset has greatly contributed to the experiments, making the predicted values more realistic, precise, and compelling. Author preprocessed the data by setting start and end dates, filtering and selecting the data, and removing unnecessary sequences and values. Additionally, a random selection was made for company samples. Table 1 show cases some instances from the dataset.

Table 1. Examples from the CAC40 dataset

	Name	Date	Open	Closing	Daily High	Daily Low	Volume
0	Accor	2020/4/3	22.99	23.4	23.4	22.99	67
1	Accor	2020/4/2	23.91	22.99	23.91	22.99	250
2	Accor	2020/4/1	24.1	23.83	24.1	23.83	37
3	Accor	2020/3/31	25.04	25	25.24	24.99	336
4	Accor	2020/3/30	26.5	25.02	26.5	24.99	415
5	Accor	2020/3/27	29.05	26.34	29.17	26.34	57
6	Accor	2020/3/26	26.39	26.39	26.39	26.39	150

2.2. Proposed Approach

The core methodology proposed in this paper for stock prediction involves the use of machine learning and LSTM models. This approach entails first collecting data such as stock prices, preprocessing the data to transform it into sequential data suitable for LSTM modeling, incorporating features like opening and closing prices from the CAC40 dataset, and then initiating simulated training with an appropriate loss function and optimizer. The model learns patterns from historical data, makes predictions, and employs these predictions to establish trading methods and strategies with the ultimate goal of assisting investors in maximizing returns. Figure 1 below illustrates the core structure of the system.

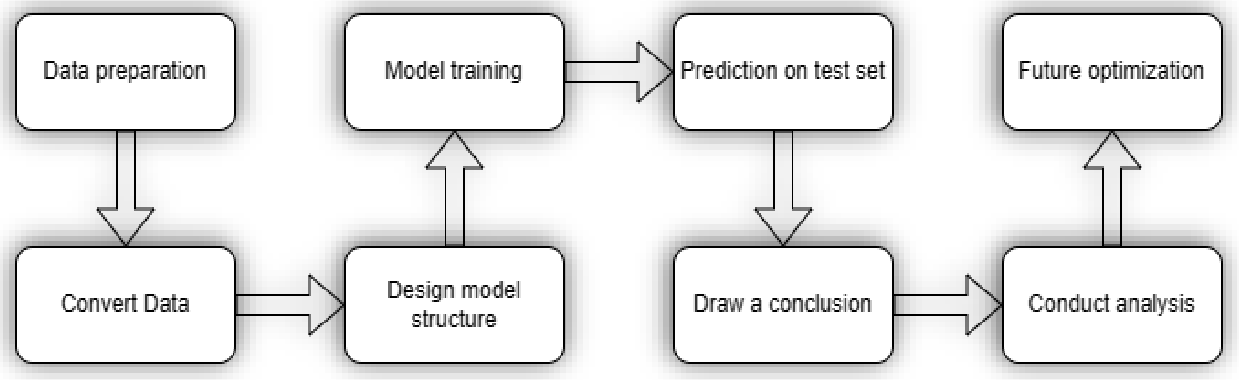

Figure 1. The pipeline of the model

2.2.1. LSTM model

LSTM, which stands for Long Short-Term Memory, is a type of temporal recursive neural network that plays a crucial role in handling and predicting significant events in time series data with relatively long intervals and delays. One key difference between LSTM and Recurrent Neural Networks (RNN) lies in the inclusion of a cell used for decision-making within the algorithm. The cell contains three gates known as the input gate, forget gate, and output gate. Data from the dataset enters the LSTM network and is assessed based on specific rules to determine its utility. Data that does not align with the rules is forgotten through the forget gate, while the relevant data is selected and outputted. This method of having two possible outputs for a single input resolves the enduring issue of repetitive computations in large datasets within neural networks, thereby improving prediction efficiency and accuracy. Over time and through empirical evidence, LSTM has proven to be a highly effective technique for addressing the computational prediction challenges posed by large datasets. Post-training, it can adeptly handle the diverse range of issues encountered in predictions. Figure 2 shows how LSTM works.

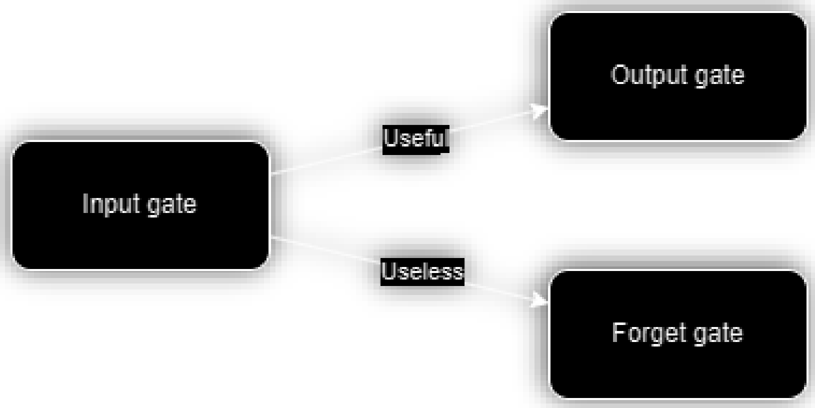

Figure 2. The principle of the LSTM

2.2.2. Loss Function

Due to the theme of deep learning in this study, selecting a suitable and optimal loss function becomes particularly crucial. The chosen loss function for this research is Root Mean Square Error (RMSE), which is a commonly used loss function in training neural networks and is well-suited for stock price prediction tasks. RMSE is capable of performing continuous numerical prediction tasks in stock price forecasting, effectively capturing fluctuations in the financial market and making adjustments accordingly. Minimizing RMSE allows for gradual parameter adjustments during training to better fit the data. As a convex function, it enables the identification of optimal solutions during training, facilitating the enhancement of stability and convergence speed in the optimization process. A smaller RMSE value indicates higher prediction accuracy and a closer fit between the model and actual values. These points underscore some advantages of RMSE, leading the study to consider it as the best loss function for addressing stock price prediction challenges, thereby driving more accurate forecasts and increased reliability.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2 \tag{1}$$

$$RMSE = \sqrt{MSE} \tag{2}$$

The formula above represents the RMSE loss function, which calculates the Mean Squared Error (MSE) first and then derives the RMSE value by taking the square root. In this formula, m denotes the number of samples summed using the summation symbol. Here, y_i represents the actual label values, and $\hat{y}_i$ denotes the values predicted by the model. Through this equation, the sum of squares of the differences between the actual label values and the model's predicted values for each sample is computed. Subsequently, the average of all samples is calculated to determine the overall MSE value, followed by taking the square root to obtain the RMSE value. By minimizing this loss function, this study aims to bring the model closer to true values and data distribution, thereby enhancing the accuracy and reliability of the trained model in predictions.

2.2.3. Implementation Details

When using models, selecting the appropriate optimizer is also a key aspect of performing deep learning tasks. In this task, this study has chosen the Adam optimizer, which is an optimization algorithm that combines adaptive learning rates and momentum. It inherits the advantages of both stochastic gradient descent and momentum optimization. Adam can rapidly converge to local optima, effectively adjust and improve learning rates, handle large datasets with ease, and is relatively user-friendly as it requires minimal manual parameter tweaking due to its default settings. Its stability and speed make it the preferred choice, serving as the optimizer for this task, minimizing the loss function and training neural network models accordingly.

3. Results and Discussion

This study delved into the analysis of stocks and the historical practice of stock prediction, visualized the CAC40 dataset, and employed an LSTM model for stock forecasting. Subsequently, the acquired dataset underwent data preprocessing to eliminate erroneous and unnecessary data, enhancing the model's effectiveness and authenticity. The LSTM model was then trained, followed by prediction post-training. Figure 3 visually represents the stock price fluctuations and trends of Atos Corporation in the CAC40 dataset from 2014 to 2020.

Figure 3. The closing price of Atos's stocks from 2014 to 2020

The Figure 3 above depicts the fluctuation trend of Atos Corporation's stock price from 2014 to 2020. Through this visual line chart, the study reveals a trend of nearly continuous growth from 2014 to 2018. However, after 2018, a downward trend emerged, with a minor recovery around mid-2019. Analysis based on the chart indicates changes in Atos Corporation's development and the broader shift in the economic landscape. It is evident that Atos Corporation experienced robust expansion from 2014 onwards. Historically, 2018 marked the 10th anniversary of the global financial crisis, where new features and heightened competitiveness emerged in the global economy due to factors such as global trade tensions, tightening monetary policies by the Federal Reserve, dollar appreciation, and tightening global dollar liquidity, leading to a decline in the equilibrium of global economic expansion post-2018. These factors contributed to Atos Corporation's economic downturn after 2018. Through data visualization, this study successfully concluded an economic analysis of Atos Corporation from 2014 to 2020. Figure 4 below uses a candlestick chart to provide a more detailed and intuitive reflection of Atos's stock price situation.

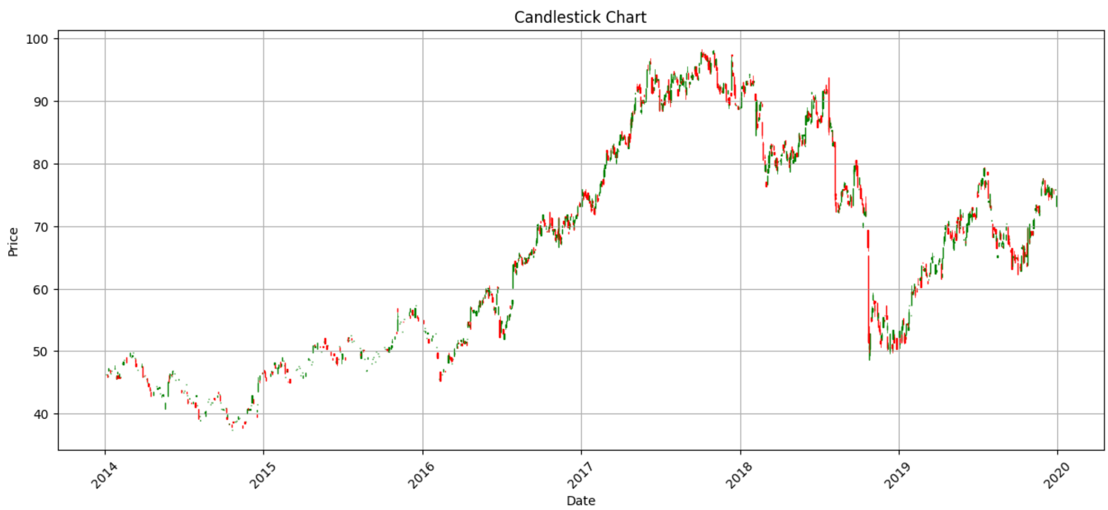

Figure 4. Presenting the same data by using candlestick chart

As shown in Figure 4, a candlestick chart can offer a more intuitive reflection of Atos Corporation's stock performance from 2014 to 2020 compared to a line chart. While its representation is more intricate than a line graph, it can depict intense market dynamics and even unveil the effective factors influencing the economy behind the scenes. Next, the author generated a comparison between a 30-day moving average line chart and a closing price line chart for Atos Corporation using the same dataset. The 30-day moving average line chart can illustrate the "life" of Atos Corporation's stock price, aiding in determining market manipulation, the presence of institutional investors, buying or selling trends, and assessing the strength or weakness of its trajectory to establish a standard. Whenever the closing price crosses below the 30-day moving average line, the stock price tends to decline, and conversely, it rises. Figure 5 is the figure that presents a comparison between the closing price and the 30-day moving average line.

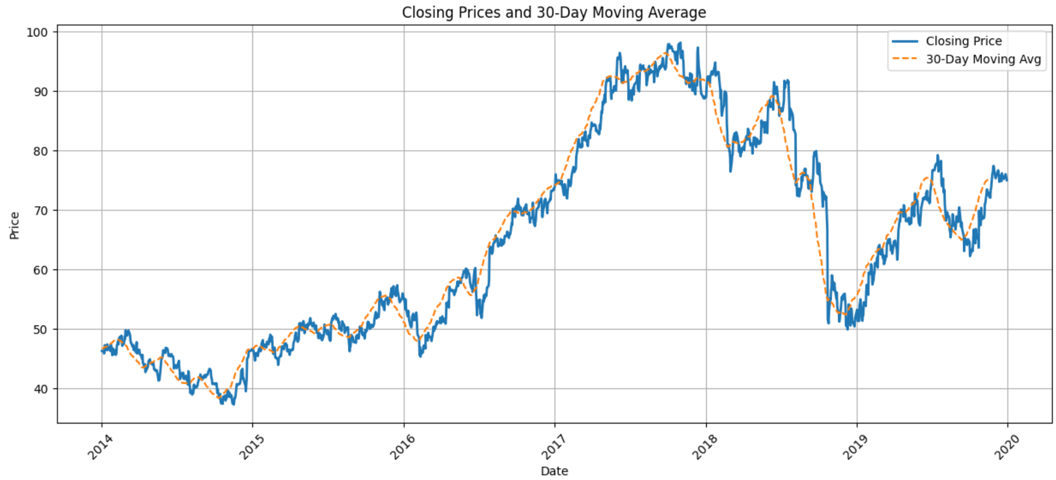

Figure 5. Closing prices and the 30-day moving average of Atos's stocks

After completing the comparison between the closing price and 30-day moving average of Atos Corporation, researchers magnified the timeline to analyze the monthly data visualization, creating a line graph with months on the x-axis and the average stock closing price on the y-axis. It can be observed that the stock closing price reaches its peak around mid-year, approximately from May to August. This indicates that Atos Corporation is more popular and sought after in the mid-year period, drawing interest from the market. This indirectly informs investors that spring, around mid-year, is the optimal time to purchase Atos Corporation stocks to maximize returns. Refer to Figure 6 for further details.

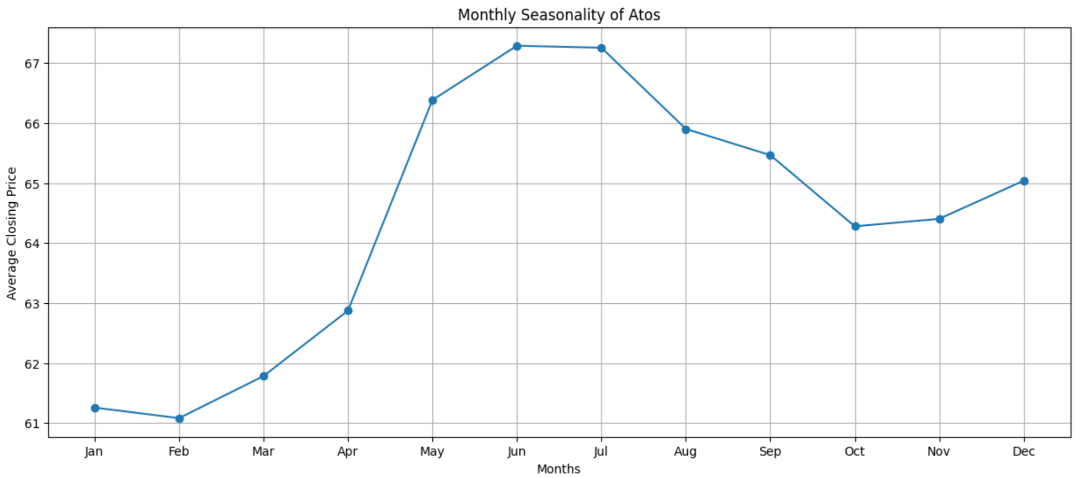

Figure 6. Monthly seasonality of the Atos's stocks

The researchers proceeded to forecast Atos stock prices by training an LSTM model. They divided the dataset into a test set and a training set. The researchers trained the model using the training set, allowing the LSTM to adjust internal parameters to minimize errors on the training data, thus enhancing its capability for future data predictions. The researchers evaluated the model's performance using the test set, comparing predicted values with actual values. The RMSE for the training set was 71.4496764574677, and when compared with the dataset, the proximity indicated a good fit for the training set. The RMSE for the test set was 43.513433753028245, demonstrating close alignment with actual values, showcasing the model's precise predictive ability meeting the research requirements. Figure 7 illustrates a comparative chart showing Atos company's stock prices between the test set, the training set, and the actual data.

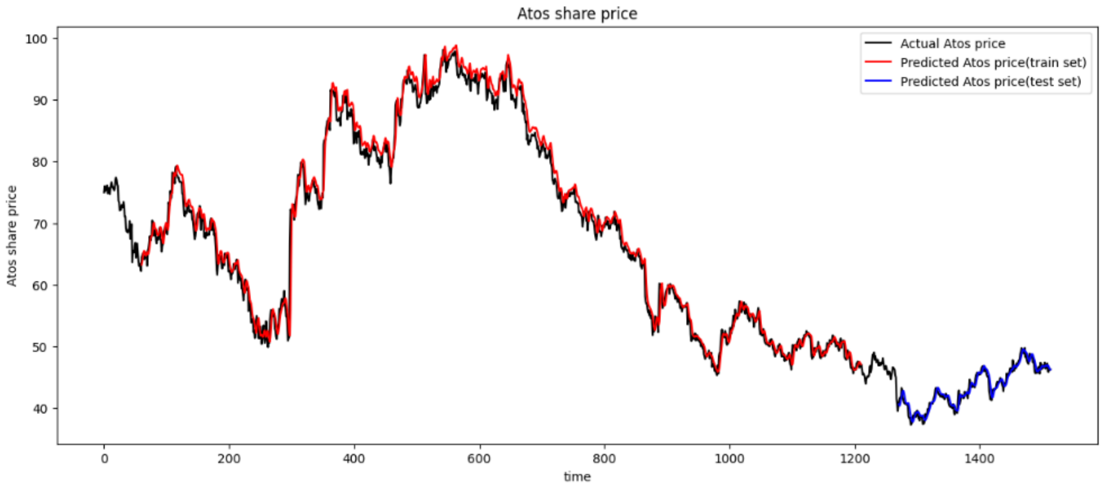

Figure 7. Actual price, train set, test set

Upon completing the training of the LSTM model, the researchers believe they can proceed with forecasting Atos company. The trained LSTM model is exceedingly accurate. Therefore, predictions for Atos company's stock prices over the next ten days were made. Based on Figure 8, an overall upward trend is observed, indicating that it is a favorable period to buy Atos company stocks. The following Figure 8 shows the projected stock price trend for Atos company over the next ten days.

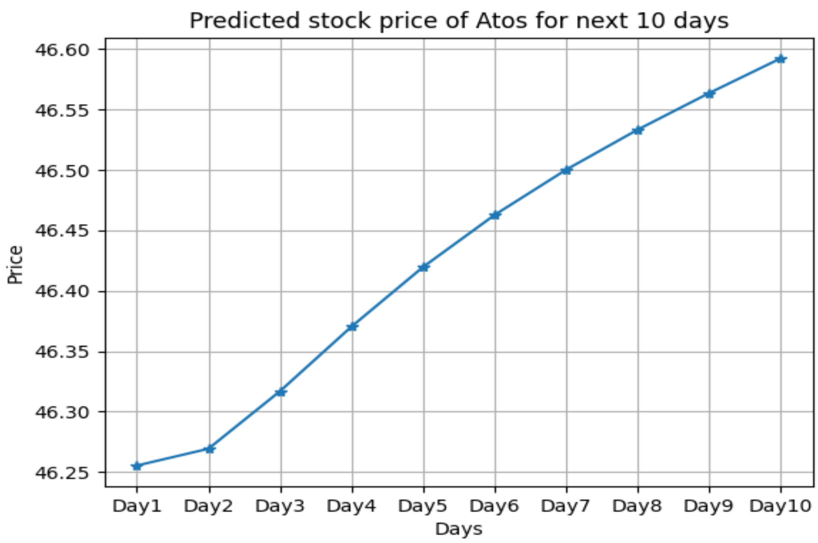

Figure 8. The stock trend of Atos company in the next ten days

In summary, the research involved visualizing the dataset for various purposes to enhance the analysis of Atos company's stocks. Following the analysis, the LSTM model was trained and utilized for stock price predictions. The forecasted results indicate that the fundamental objective of the study - to empower investors to make informed decisions, maximize profits, and provide valuable insights - has been achieved. Taking Atos company as an example, purchasing its stocks over the next ten days would result in a non-loss position, affirming a positive outlook for investors.

4. Conclusion

The LSTM model utilized in this study has made significant contributions, and the calculation of RMSE has advanced the research. The use of the LSTM model effectively predicted the research goals and demonstrated advantages over traditional forecasting methods when dealing with extensive data. It offers higher accuracy in forecasting stock tasks, assists in providing purchase recommendations, and aids investors in maximizing profits. By integrating deep learning and machine learning, it showcases diverse fusion features to address tasks involving substantial data volumes, such as stock price predictions. In the future, the research will delve deeper into predicting tasks with even larger data volumes, utilizing larger training and test sets to refine the model's accuracy. Considering its limitations, the research will further explore comparative analyses among multiple stocks to expand beyond singular stock purchase advice and enable investors to select from various companies, maximizing returns. The next focus of the research will encompass multi-stock predictions and data visualization, aimed at enhancing broader stock price predictions for increased practicality, accuracy, and authenticity.

References

[1] Z. Javad, and M.M. Rounaghi. Application of Artificial Neural Network Models and Principal Component Analysis Method in Predicting Stock Prices on Tehran Stock Exchange. Physica A: Statistical Mechanics and Its Applications, 438, 2015, pp. 178–87.

[2] P. Jigar, et al. Predicting Stock Market Index Using Fusion of Machine Learning Techniques. Expert Systems with Applications, 42(4), 2015, pp. 2162–72.

[3] Y. FM, et al. A Novel Hybrid Stock Selection Method with Stock Prediction. APPLIED SOFT COMPUTING, 80, 2019, pp. 820–31.

[4] Information on: https://api.semanticscholar.org/CorpusID:210836684.

[5] N. Mojtaba, et al. Deep Learning for Stock Market Prediction. 2020.

[6] G. Aditya, and B. Dhingra. Stock Market Prediction Using Hidden Markov Models. Engineering & Systems, 2012.

[7] M. XH, and H. Lin. Predicting Stock Market Crises Using Stock Index Derivatives: Evidence from China. EMERGING MARKETS FINANCE AND TRADE, 60(3), 2024, pp. 576–97.

[8] Wang. YF. Predicting Stock Price Using Fuzzy Grey Prediction System. EXPERT SYSTEMS WITH APPLICATIONS, 22(1), 2002, pp. 33–38.

[9] M. Ananthi, and K. Vijayakumar. Stock Market Analysis Using Candlestick Regression and Market Trend Prediction (CKRM). JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING, 12(5), 2021, pp. 4819–26.

[10] T. LW, et al. Stock Price Prediction Based on LSTM and LightGBM Hybrid Model. JOURNAL OF SUPERCOMPUTING, 78(9), 2022, pp. 11768–93.

[11] Information on: https://www.kaggle.com/datasets/bryanb/cac40-stocks-dataset/data.