Analysis of Different Machine Learning Models for Stock Price Prediction

Dhanaraj Patil - 820905822, Mohit Agarwal - 825852775

Abstract

In stock market investing, accurate predictions are crucial. Predicting stock values is a difficult and dangerous task. If forecasts are not accurate, there is a considerable chance of losing money in a matter of seconds. We implemented two such models to predict stock prices in this project. The goal is to assess the accuracy of predictions for different machine learning and deep learning models in order to estimate future pricing for a certain company. While it is practically difficult to anticipate the exact value of a stock, we may use these models to examine the trend of stock prices in the past to determine future actions.

Introduction

The stock market [1] is one of the most popular public markets for individuals and businesses to invest in long and short term. All companies with shares in the market gain capital, which they can use to fund and develop their operations. The stock market also allows investors to participate in the profits of publicly traded corporations by purchasing stocks. Company shares were originally issued on paper, and investors would trade them back and forth using the papers. Until the creation of the London Stock Exchange [1], there were no regulated exchanges. Since then, the process of trading stocks has developed. Recently, several trading algorithms that use machine learning methods such as artificial neural networks have outperformed other conventional methods.

From various people in our lives, we have all heard about the ups and downs of stock investing and market volatility. We, like many others, took a chance and invested in the stock market. While our investment experiences were highly varied, the trend of stock price gain and fall was common. To invest successfully in the stock market, you must first learn how to trade, the stock market's pattern, risk analysis, and so on. While I (Dhanaraj) have had stable stock market investments for a couple of years now, Mohit on the other hand was unfamiliar with the concept of stock market investing. He was urged by experienced investors to invest in a certain stock that was profitable at the time. Due to his lack of knowledge, He invested without considering other considerations and lost money within a few weeks when the stock price dropped dramatically. Our own stock market experiences inspired us to create a TradeBot that would primarily assist newbies like us in avoiding a financial loss in stock investments. The TradeBot will examine historical stock price trends and forecast the stock's future value. Because there are so many models on the market, we chose models that forecast using linear regression, deep learning, and time series. We are not inventing a new idea by executing this project; rather, the purpose of our project is to gather in-depth knowledge of the AI area working with a huge data set.

Related Work

Fama, E. F. (1970) proposed the Efficient Market Hypothe- sis (EMH), while Horne, J. C., and Parker, G. G. (1967) proposed the Random Walk theory in early stock market prediction research. These ideas claimed that market prices are influenced by information other than historical prices and hence cannot be predicted [2], but current research has proven that stock market price movement may be predicted to some extent. This forecast is based on two different forms of financial analysis:

- Fundamental Analysis

Fundamental analysis is focused on the company's health, as well as qualitative and quantitative factors such as interest rates, revenues, and price to earnings, among others. The goal of this examination is to look at the company's long-term viability and strength as a long-term investment.

- Technical Analysis

Traders who use technical analysis look at historical prices and charts and use time as a critical aspect in their predictions. Technical analysis relies on three basic elements: stock price movement, which might appear random at times, historical trends that are thought to reoccur overtime, and all pertinent information about a stock.

Various machine learning approaches have been applied to predict future stock prices in recent studies. While no prediction can be made with 100 percent accuracy, we may rely on these forecasts to deliver a price decrease or increase. There are numerous techniques of predicting stock prices for traders to make the best trading decisions. In economic domains, technical indicators such as moving averages have been developed to identify market patterns, and many computer science investigations are now aimed at predicting future prices, mostly using AI and Machine Learning technology.

While machine learning methods such as Linear Regression and K-nearest neighbors might forecast stock prices for the following day, it was revealed that utilizing time series, we could predict future stock prices. This was made possible by algorithms such as ARIMA, LSTM, and Prophet. These models primarily focus on extracting patterns from historical data in order to forecast future values.

Methods

For this project, we implemented two models for predicting future stock price values for a certain company. Our research is based on stock prices of different companies.

We obtained stock prices from CAC40 Stocks Dataset [7] to give data for each model. This project utilized data from the previous five years.

To compare the accuracy of predictions of models, we have selected the following models:

1. LSTM

Long Short Term Memory networks, often known as "LSTMs," are a type of recurrent Neural Network capable of learning long-term dependencies. Their default behavior is to remember information for a long time. LSTMs, like all recurrent neural networks, contain a chain-like structure of repeated neural network modules. These repeating modules in LSTMs have varied architectures. It comprises four neural network layers that interact in a unique fashion, rather than a single neural network layer. The memory cell introduced by LSTM allows for long-term reliance between temporal delays. Memory cells in the RNN replace hidden layer neurons and filter information through the gate structure to preserve and update memory cell states. Input gate, forget gate, and output gate are all part of the gate structure. Because they can store prior knowledge, LSTMs are highly powerful in sequence prediction issues like stock price predictions. This is critical in stock price prediction since a stock's previous price is significant in predicting its future price.

We followed the procedures below to create the LSTM model for stock price prediction.

- Read and examine the data obtained from Kaggle
- Normalize the dataset to aid the LSTM algorithm in locating the local/global minimum efficiency.
- The data is converted to time-series and supervised learning problems, which means the dataset is converted into a three-dimensional array as required by the LSTM algorithm.
- Creating an LSTM model and feeding it the training data as input. The model learns from this information and examines prior stock prices.
- The output of the model using the testing dataset is plotted on a graph once it has been trained.

2. ARIMA

The Autoregressive Integrated Moving Average (ARIMA) model is a well-known and commonly used forecasting method for time series prediction. ARIMA models are derived from statistical models. Predictions can be done from two angles, according to the literature: statistical and artificial intelligence techniques. ARIMA models are known to be more robust and efficient than even the most popular ANNs techniques in financial time series forecasting, especially short-term prediction.

Using the following equation, the future value is determined using a linear combination of past values and past errors:

$$\left(1-\sum_{i=1}^{p'}\alpha_i L^i\right) X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$

where L is the lag operator, αi is the autoregressive part of the model's parameters, θi is the moving average component's parameters, and εt is the error terms. Model identification, parameter estimation, and diagnostic checking are the processes in developing an ARIMA predictive model.

Root Mean Squared Error [5]

The root-mean-square error (RMSE) is a commonly used measure of the variations between predicted and observed values (sample or population values) by a model or estimator.

The RMSE formula is as below, Where f = forecast (expected values or unknown results) and o = Observed values (known results)

$$RMSE = \sqrt{\overline{(f-o)^2}}$$

If the RMSE number is 0, the model's prediction is considered accurate. A value of zero indicates that the model's predicted values precisely match the test's actual values. The square root of the difference between predicted values and test dataset is determined in our application using the NumPy module.

Implementation

Libraries used :

1. NumPy
2. Pandas
3. Sklearn
4. Statsmodels
5. Keras
6. Matplotlib
7. Streamlit for graphs

Procedure to implement and predict values:

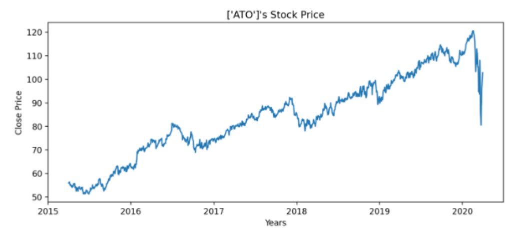
1. Read historical data of stock prices from a CSV file.
2. Plot and visualize the current stock prices.
3. Divide the input dataset into 2 parts, training and testing
4. Build the program model and feed it the values of the training model.
5. Predict future values using the trained model.
6. Using the predicted values and testing dataset values, calculate RMSE value.
7. To visualize the performance, plot the training dataset, testing dataset and predicted values to the graph

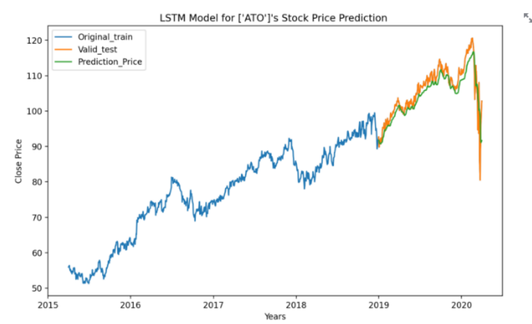
Notes:

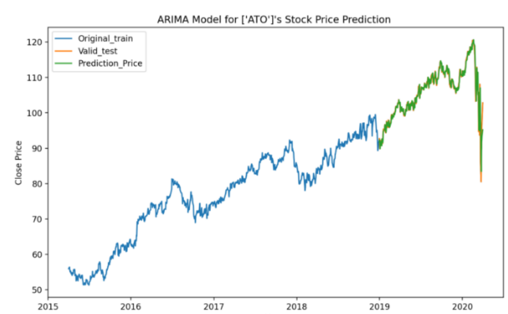
1. Historical data of stock prices is taken from Kaggle [7] website. The models are trained and tested using data from the previous five years.
2. We considered the stock values of ATOS, Renault, L'Oréal, Schneider Electric for this project.
3. The training and testing datasets are the same size for all models. Training accounts for 75% of the input dataset, whereas testing accounts for 25%.
4. Calculating the square root of variations in values between predicted values and testing the dataset yields the RMSE value for each model.

Results and Analysis

We evaluated the outputs of both the models for the stock prices of ATOS in order to compare their performances in terms of stock price forecast accuracy. We generated the following outputs for the models using data from 2015 to 2020.







We used the RMSE factors of each model in addition to the graphical representations of the results. The following RMSE factors were calculated based on the data.

Model Name	**RMSE Factor**
LSTM	3.414
ARIMA	2.378

We discovered that when trained on stock prices that follow a linearly increasing pattern, the ARIMA model has the highest precision for predicted values on the test dataset among the 2 algorithms used for the comparison.

Due to pandemic, in the year 2020 market prices were particularly dramatic as a result of the ongoing issues on the world platform, which is the test data and projection dates for these models. That is why we wanted to check the prediction during these times hence we ran the ATOS stock during the period range 2015 - 200 March ignoring the pandemic days, and our models performed really well.

The RMSE values for the same are shown in the table below.

Date Range 2015-April --- 2020-March

Model Name	**RMSE Factor**
LSTM	3.019
ARIMA	1.928

Conclusion

We created a stock price prediction system with different prediction models for this project to assess the accuracy of their anticipated values. For this we ran the model on different companies from the CAC40 Stocks Dataset during the last five years. On this historical data, the models LSTM and ARIMA were trained, and predictions with RMSE values were calculated.

Based on the results, we found that the time series-based algorithms performed better than regression algorithms. For all the companies below, the ARIMA model was the most accurate in stock price projections.

Stock	LSTM	ARIMA
ATOS	3.414	2.378
Renault	2.267	1.001
L'Oréal	0.426	0.339
Schneider Electric	1.800	0.731

ARIMA outperformed the other model and was able to predict prices that were very close to the actual prices. Because the historical price behavior was comparable to the train data set, it appears that predicted values are quite close to the real values for Schneider Electric. There was also no dramatic price increase. And it also predicted relatively bad RMSE value for ATOS since the test data showed the ATOS price fluctuated dramatically during the pandemic.

As a result, we can conclude that these models perform effectively when historical trends (price fluctuations) are similar and there are no dramatic changes in the price pattern.

While Stock prices are affected by a variety of unknown factors such as election results, rumors, political effects, quarterly performances and so on, machine learning models cannot anticipate these factors. We assumed for the purposes of this project's implementation that the historical data had already included the influence of these unknown elements into their pattern.

Contribution

We both worked together on this project. We together decided the approach, analysis, report and other research related to the project.

References

[1] “Stock Markets and how they work”, Corporate Finance Institute, https://www.investopedia.com/terms/s/stockmarket.asp

[2] Mokalled, Wassim El-Hajj Mariam and M. Jaber. “Automated Stock Price Prediction Using Machine Learning.” (2019)

[3] Adebiyi, Ayodele & Adewumi, Aderemi & Ayo, Charles. (2014). “Stock price prediction using the ARIMA model”, Pro- ceedings - UKSim-AMSS 16th International Conference on Computer Modeling and Simulation, UKSim 2014. 10.1109/UKSim.2014.67.

[4] Selvin, Sreelekshmy et al. “Stock price prediction using LSTM, RNN and CNN-sliding window model.” 2017 International Con- ference on Advances in Computing, Communications, and Infor- matics (ICACCI) (2017): 1643-1647.

[5] Root Mean Square Deviation, Wikipedia, *https://en.wikipedia.org/w/index.php?title=Root-mean-square_deviation&action=history*

[6] R. Sathya, Prateek Kulkarni, Momin Nawaf Khalil, Shishir Chandra Nigam, Stock Price Prediction using Reinforcement Learning and Feature Extraction. In International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume- 8 Issue-6, March 2020

[7]https://www.kaggle.com/datasets/bryanb/cac40-stocks-dataset