Analysis of Machine Learning Models for Predicting Stock Prices

Group Members:

Aakash Joshi (825859093)
Dhanaraj Pati (820905833)
Dhruv Makati (825864917)
Himani Keskar (825026807)
Mansi Vyas (825866373)

Abstract

In stock market investing, accurate predictions are crucial. Stock price forecasting is a challenging and dangerous task. There is a considerable chance of losing money in a couple of seconds if projections are inaccurate. In this study, we used two of these models to forecast stock prices. The objective is to evaluate the reliability of forecasts for various deep learning and machine learning models in order to project future prices for a certain company. Although it is nearly impossible to predict a company's precise value, we may use these models to look at the historical pattern in stock prices to guide our future decisions.

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.

We have all heard about the ups and downs of stock investment and market volatility from various people in our lives. Like many others, we gambled and made stock market investments. Despite the wide range of our financial experiences, the pattern of rising and falling stock prices was typical. We must first understand how to trade, the patterns of the stock market, risk analysis, and other skills if we want to make money investing in stocks. While a couple of team members have had stable stock market investments for a few years now, others on the other hand were unfamiliar with the concept of stock market investing. While some of us urged by experienced investors to invest in a certain stock that was profitable at the time. Due to our lack of knowledge, We made investments without taking other factors into account and lost money within a few weeks when the stock price dropped dramatically. Our own stock market experiences motivated us to create a TradeBot that would help novices like us in avoiding a financial loss in stock investments. The future value of the stock will be predicted by the TradeBot by looking at previous stock price trends. We selected models that employ time series, deep learning, and linear regression to forecast because there are so many models available. The goal of our research is to get in-depth knowledge of the AI field while working with a sizable data collection, not to come up with a novel idea.

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 theories asserted that market prices are affected by information other than past prices and cannot, therefore, be forecast [2], but recent research has shown that stock market price movement may, in fact, be somewhat predicted. This prediction is supported by two types of financial analysis:

- Fundamental Analysis

Fundamental analysis is focused on a company's health as well as qualitative and quantitative elements like interest rates, sales, and price to earnings, among others. This analysis aims to evaluate the company's long-term strength and feasibility as an investment.

- Technical Analysis

Technical analysts look at previous prices and charts and make predictions using time as a key factor. Three fundamental components are used in technical analysis: stock price movement, which might occasionally seem random, historical trends that are believed to recur over time, and all relevant information about a stock.

In recent research, a variety of machine learning techniques have been used to forecast stock values. We may rely on these projections to bring about a price fall or increase even though no prediction can ever be made with 100% precision. For traders to make the greatest trading choices, there are several methods for forecasting stock prices. Many computer science projects are now targeted at forecasting future prices, particularly utilizing AI and Machine Learning technologies. Technical indicators like moving averages have been established in the economic domains to discover market trends.

While machine learning techniques like linear regression and K-nearest neighbors may anticipate stock prices for the following day, it has been shown that using time series, we may estimate future stock prices. Algorithms like LSTM, ARIMA, and Prophet made this feasible. In order to anticipate future values, these models typically concentrate on identifying patterns from previous data.

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 uses dataset from 2010 to 2021.

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

1. LSTM

LSTMs, or long short term memory networks, are a form of recurrent neural network that can learn long-term dependencies. They naturally have a long-term memory for knowledge. Like all recurrent neural networks, LSTMs have a structure resembling a chain made up of repeated neural network modules. The topologies of these recurring modules in LSTMs vary. Instead of a single neural network layer, it consists of four neural network layers that interact in a special way. Long-term dependency between temporal delays is possible because to the memory cell that LSTM created. In order to retain and update memory cell states, memory cells in the RNN take the role of hidden layer neurons and filter input through the gate structure. The gate structure includes the input gate, forget gate, and output gate. LSTMs are extremely effective in sequence prediction problems like stock price forecasts because they can store past information. This is crucial for stock price forecasting since a stock's historical price influences future price predictions.

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

- Reading and analyzing the Kaggle data
- Normalize the dataset to aid the LSTM algorithm in locating the local/global minimum efficiency.
- The dataset is transformed into time-series and supervised learning problems, which means it is transformed into a three-dimensional array in accordance with the LSTM algorithm's specifications.
- Making an LSTM model, then feeding it the training set of data. This data is used to train the model, which also looks at previous stock values.
- Once the model has been trained, the results using the testing dataset are shown on a graph.

2. ARIMA

For the prediction of time series, the Autoregressive Integrated Moving Average (ARIMA) model is a well-known and often used forecasting technique. Statistical models are used to create ARIMA models. According to the literature, predictions can be made using statistical and artificial intelligence approaches. In financial time series forecasting, especially short-term prediction, ARIMA models are recognized to be more reliable and effective than even the most widely used ANNs approaches.

The future value is calculated using a linear combination of previous values and past mistakes using the following equation:

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

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.

Approach

Libraries used :

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

Procedure to implement and predict values:

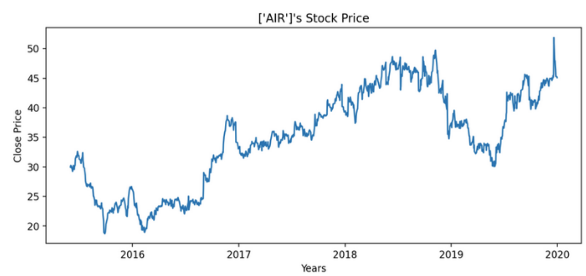
1. Read historical stock price information from a CSV file.
2. Visualize and plot the current stock prices.
3. Separate the input dataset into training and testing sections.
4. Create the program model and feed it the training model's values.
5. Utilize the learned model to make future value predictions.
6. Calculate the RMSE value using the values from the testing dataset and the anticipated values.
7. To visualize the performance, plot the training dataset, testing dataset and predicted values to the graph

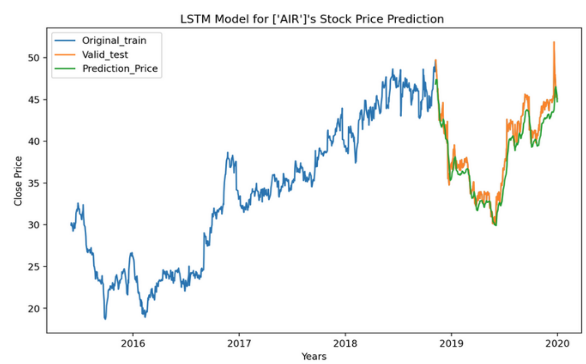
Notes:

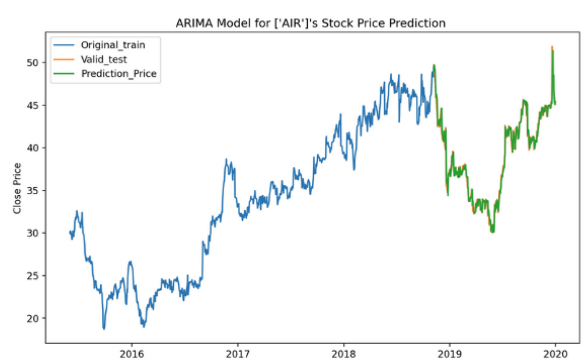
1. Stock price historical information is sourced from the Kaggle [7] website. Data from 2010 to 2021 is used to train and evaluate the models.
2. For this study, we took into account the stock prices of AIR, ATOS, AXA, and Accor.
3. For each model, the size of the training and testing datasets is the same. 25% of the input dataset is used for testing, while 75% is used for training.
4. The RMSE value for each model is obtained by taking the square root of the differences between predicted values and testing the dataset.

Evaluation/Analysis of Results

In order to compare the performances of the two models in terms of stock price forecast accuracy, we assessed the outputs of both models for the prices of AIR. Using information from 2015 to 2020, we produced the following results for the models.







Along with the results' graphical representations, we also utilized each model's RMSE factors.

Based on the data, the following RMSE values were determined for AIR.

Model Name	**RMSE Factor**
LSTM	1.590
ARIMA	0.919

We found that of the 2 methods utilized for comparison, the ARIMA model had the greatest precision for projected values on the test dataset when trained on stock prices that follow a linearly rising trend.

The test data and projected dates for these models show that market prices were particularly dramatic in 2020 as a result of the pandemic and the continued problems on the global platform. We ran the ATOS stock throughout the period range of 2015 - 2020 January ignoring the pandemic days since we wanted to validate the prediction during these periods, and our models worked pretty well.

Conclusion

For this project, we developed a stock price prediction system using various prediction models to evaluate the precision of their predicted values. For this, we ran the model over the previous five years on a variety of businesses from the CAC40 Stocks Dataset. The models LSTM and ARIMA were trained using this historical data, and forecasts with RMSE values were computed as shown below.

Stock	LSTM	ARIMA
AIR	1.590	0.919
AXA	0.251	0.175
Accor	1.117	0.721
ATOS	1.429	0.993

Based on the findings, we discovered that time series-based algorithms outperformed regression techniques. The ARIMA model's stock price estimates were the most accurate for all the firms listed below.

The ARIMA model outperformed the other one and was able to accurately anticipate prices that were quite close to the current prices. Given that the historical pricing behavior and the train data set were comparable, it seems that AXA's projected values were very close to the actual values. Additionally, given that the test data revealed that the price of AIR changed significantly during the pandemic, it also projected a somewhat poor RMSE value for AIR.

As a result, We may conclude that these models work well when historical trends (price fluctuations) are similar and there are no dramatic changes in the price changes.

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.

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