Shaili Sharma. 2020, 8:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Stock Prediction using Machine Learning

Shaili Sharma
Associate Software Engineer, Universal Softlab, shailisharma0212@gmail.com

***Abstract*- In the world of business, the stock market, often known as the stock exchange, is one of the most complicated and sophisticated methods of doing business. When addressed with discipline, the stock market may be very volatile, but it can also be one of the most successful techniques of making large gains. A variety of entities, including small enterprises, brokerage firms, and the financial industry, depend on this organization to earn money and disperse risks. By using open-source libraries and current approaches, this effort aims to develop machine learning models in a WebApp that can be used to anticipate stock values in the future for the purpose of trading, in order to help make this volatile kind of commerce a bit more predictable. In order to avoid using the standard technique and from receiving results that are just dependent on statistics, a system that incorporates a text-based machine learning model as well as pattern recognition is being developed. This project's goal is to provide a platform for small and amateur traders to anticipate future stock market values using current stock prediction models that take into account factors such as recent news articles, recent stock volume, and the previous close, among other things.**

Keywords—Stock market prediction, machine learning, Time series, Sentimental Analysis.

I. INTRODUCTION

The stock market is one of the oldest means of earning money from firms that sell a portion of themselves on this platform. An average individual could trade stocks, make investments, and earn money from companies that sold a portion of themselves on this platform. In the modern world, the stock market serves as a platform for practically all significant economic transactions, with the stock value fluctuating at a dynamic pace that is determined by market equilibrium. If implemented properly, this technique has the potential to be a profitable investment strategy.

Predicting the value of a stock provides great profit potential, which serves as a significant motivator for further study in this field. Earnings may be generated by knowing the value of a stock for even a fraction of one second. The same is true in the repeating context: a probabilistically right forecast may be quite beneficial in this situation. As a result of the allure of discovering a solution, researchers in both business and academia have been working hard to find a means to get beyond issues such as volatility, seasonality, and dependency on time, economics, and the rest of the market. Prices and liquidity on the platform, on the other hand, are very unpredictable, which is where technology might be of assistance.

In the past, one of the most active research topics was the forecast of the stock market. To create money and disperse risk, small businesses, brokerage firms, and banking sectors all depend on the stock market to do business and make a profit. As a consequence, anticipating the value of a company's stock is of paramount importance. If we utilize crowd computing to graph the stock exchange price, we will return an approximate response to a real-life graph, but it will be a very long process because of the large number of participants. Recent breakthroughs in deep learning have enabled trading algorithms to become increasingly accurate in their predictions of stock price fluctuations. Unfortunately, there is a significant gap between the development of this discovery and its application in the actual world. Furthermore, these cutting-edge technologies are seldom used to the advantage of small-scale sellers.

The goal of "StockClue" is to develop an interactive web-app with effective machine learning models that will assist many small-scale investors in making long-term investments. The app will be a fully-functional website with all stock information, such as a candlestick graph, open price, closing price, volume, and pertinent news, among other things.

II.RELATED WORKS

According to Lei Shei et al. [1], the architecture of the neural network model was based on data from news sources and Twitter tweets]. To connect each item of news to the right stocks, each company maintains a list of keywords that may be searched (e.g., Apple: AAPL, AAPL.O, APPLE, AAPL.N, Apple Inc, etc.). Using the Twitter API, the stock-related tweets were retrieved by searching for hashtags associated with the company in the tweet text. With this approach, the goal is to estimate a stock price y that is as near as possible to the company's actual stock price y.

DeepClue is built with the help of DyNet v1.0, a neural network software library that is specifically designed for natural language processing applications. From 2006 to 2015, they looked at stocks in the S&P 500 index on the New York Stock Exchange. Their historical prices are obtained from Yahoo Finance, while their financial news is obtained from Reuters and Bloomberg, among other sources.

A model created by Yangtuo Peng et al. [2] computes the closing prices of the previous five days in order to provide an input feature vector for deep neural networks. The model examines all financial periodicals for sentences that include at least one stock name or public company as a part of the subject matter. Each phrase is separated into samples and labelled with the publication date of the original article as well as the stock name associated with it. Each example consists of a collection of statements that were published on the same day and that reference the same stock or organization. Each sample is additionally labelled as either positive ("price-up") or negative ("price-down") based on the closing price the next day of the sample. The DNN is used, which contains layers that are concealed from view (each with 1024 hidden nodes). As a starting point, the historical pricing feature is employed, and other financial news-derived aspects are layered on top of it. The DNN outputs are divided into groups based on the dates of the samples included in the test set. All of the forecasts for unseen stocks are compared to the actual stock movement the next day. Data from Reuters and Bloomberg were utilized to compile this paper's financial news analysis. The historical stock security information is derived from the CRSP data repository (Center for Research in Security Prices).

According to Wasiat Khan et al. [3, they used the financial news site Business Insider for analysis, and they used the Stanford emotional analysis package of Stanford NLP, which awarded points for positive and negative terms, respectively."

Researchers Xianghui Yuan and colleagues [4] developed a model to predict the stock's excess returns for the next month. In order to get 60 characteristics that will be used as input to the model, financial reports, daily starting prices, daily closing prices, volumes, and other data from the A-share market during an eight-year period are gathered and analysed.

When labelling data, Jingyi Shen et al. [5] developed a methodology to determine the price trend by comparing the current closing price to the closing price of n trading days earlier, which they called the price trend model. The LSTM method is used for time-series prediction, which assures that the prediction model can capture both sophisticated hidden patterns and patterns that are connected to time series. There are 3558 equities from the Chinese stock market in this dataset. It also includes data obtained via an open-sourced API and data scraped from Sina Finance web pages and the SWS Research website.

In their paper [6, Guangyu Ding and colleagues suggested an associated deep recurrent neural network model with multiple inputs and multiple outputs that was based on a long short-term memory network. In the case of a stock, the related network model is capable of predicting its opening price, lowest price, and peak price all at the same time. The experimental data used in this paper are actual historical data downloaded from the Internet, the Shanghai composite index 000001, and two stocks: PetroChina (stock code 601857) on the Shanghai stock exchange and ZTE (stock code 000063) on the Shenzhen stock exchange. The experimental data used in this paper are actual historical data downloaded from the Internet, the Shanghai composite index 000001, and the other is the Shanghai composite index 000001.

Abdalraouf Hassan and colleagues [7] Using a joint CNN and RNN framework, a set of feature maps learned by a convolutional layer with long-term dependencies learned via long-short term memory is combined with an unsupervised neural language model to train initial word embeddings tuned by a deep learning network, after which the network's pre-trained parameters are used to initialise the model. This study examined how well the suggested model performed using data from the Stanford Large Movie Review dataset (IMDB) and the Stanford

Sentiment Treebank dataset (SSTb), which was compiled from Rotten Tomatoes movie reviews.

Glove was developed by Yash Sharma et al. [8, who also discuss its potential use in sentiment analysis. The word vectors produced by the Glove approach are input into an RNN, which performs binary classification using sentiment analysis (Positive and Negative Sentiments).

In [9], D V Nagarjana Devi and colleagues developed the HARN algorithm, an unsupervised learning system that leverages the fundamental structure of the sentences, domain dictionaries, and predetermined polarity to categories a given phrase into one of three categories.

III.PROBLEM STATEMENT

Because just a limited dataset is utilized for training, the accuracy of present stock market prediction models is rather low. As a consequence, the outcomes will be less accurate than they otherwise would be. There is still a need to investigate additional new features that are more predictable on a continuous basis. Despite the fact that several algorithms exist, there is no real-world application of these concepts for the benefit of the general public. Efficient algorithms should be made accessible to the public with a simple interface and accessibility.

IV.PROPOSED SYSTEM

According to the suggested technique, an interactive online platform (Web App) for stock traders will be developed so that they may use it to estimate the future value of the stock market. In addition, the Web App displays market prices, volume, and other information, as well as a prognosis for the stock in question. The objective is to develop a platform that contains a large number of efficient stock market machine learning techniques. Finance News and Stock data are two of the stock prediction criteria that may be learnt by an individual. Generally speaking, the complete Web App may be broken down into the following sections:

WebApp contains a page for user authentication where individuals may login to the app to examine the real-time status and forecast of numerous stocks.

2. If the user is logging in for the first time, he or she may also register.

A slider with listed stocks, a current stock price display, and forecasted values may all be found on the live page. After selecting desired stocks from the slider, the matching live stock values, forecasted graphs, and ML model results will be presented on a separate screen.

Third, the current news concerning the specific stock will be presented in a new bar, which the user may scroll through and also click on a link that will take them to the news piece in question.

Accessed through the cloud, the most recent news is retrieved and shown in the online application; the same content is also vectorized according to its positivity or negative factor and displayed in order to analyze its influence on the stock price. To estimate future prices, the closing price of the market is gathered and used in conjunction with the LSTM algorithm. It is then rendered visually in the web application. The XGBoost prediction model gathers the OHLC (open, high, low, close) price, as well as volume and the news vector, in order to forecast the trends for the next day. It is also shown in the online application by means of a graph.

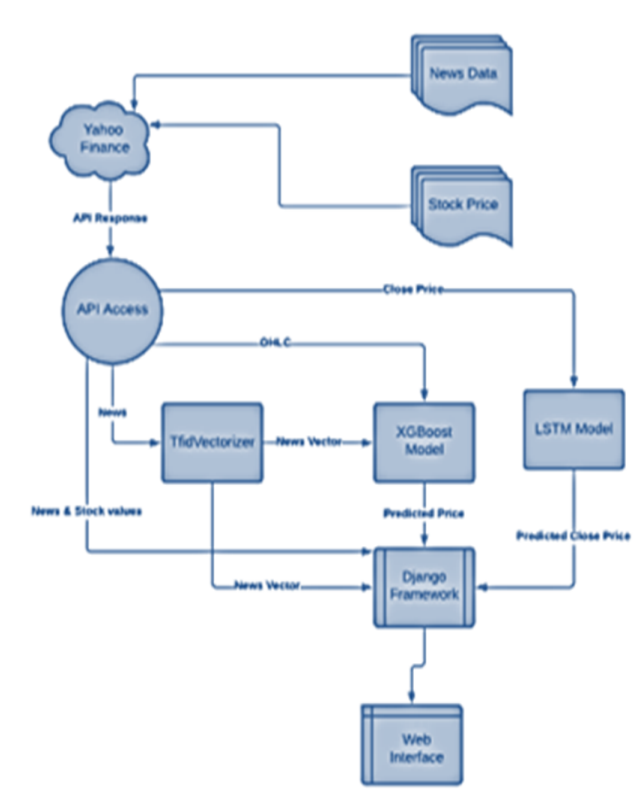

Fig 1 Stock Clue Architecture Design Complete backend design of the web app.

V.IMPLEMENTATION

The Python programming language and the Django framework are used to build the backend of the

whole project. PostgreSQL is in charge of keeping track of user data and login details. When developing a reactive web application, HTML, CSS, and JavaScript are all used efficiently. The stock charts are shown via the usage of the Chart.js module.

Live Data: Machine learning models need current data such as news, volume, and so on. The data is also required by the online application in order to show the current stock graph. As a result, the data is obtained using the Yahoo Finance API and the nsepython module.

Models based on machine learning: There are three stock prediction models in the app. Sentiment Vector: This function determines the trend of the most recent news about the stock. The LSTM algorithm determines the closing price for the following 30 days. XGBoost forecasts the stock price based on news data, volume, the previous close, the open, and other factors.

One of the most often used models is the LSTM model, which is used to forecast the closing price of a stock for the following 30 days from the present date. This model is trained using the closing price of the specific stock in order to forecast the closing patterns of the stock in future periods of time. For LSTM, the closing price of 'Infosys' during the last 20 years was used. Eighty percent of the data was utilized for training, while the remaining twenty percent was used for testing. A 99 percent accuracy rate was achieved with respect to the test data by the model. After that, the model was saved using the webapp. The dataset was obtained from the Kaggle website.

News Vector: To evaluate the most recent news data, the Tfidf Vectorizer from scikit-learn is used as an example. A vectorized value for each piece of news is returned by the model, which accepts as input the news data. It conveys both the positive and negative aspects of the news. It is then integrated into the online application. This data is retrieved from the Yahoo Finance API, which is updated in real time.

In terms of accuracy and reasonableness, the XGBoost algorithm provides the most accurate and reasonable forecast. The model takes as inputs the news vector, the open price, the close price, and the volume for the previous three days, and predicts the closing price for the next three days.

The model was trained using data collected over a five-year period. The news data is obtained by web scraping, while the quantitative data is obtained through Kaggle. The dataset was pre-processed in order to integrate the news data with the quant data based on the date of the news article. The data was utilized for training purposes 80 percent of the time, and testing was done with the remaining 20 percent. The model demonstrated accuracy of 90 percent. It is then stored via the online application.

VI.RESULTS AND COMPARISONS

An interactive web app to predict the stock market was made to help small scale traders to efficiently invest reducing risk. Our works clearly showed an increased accuracy in stock prediction compared to other similar works.

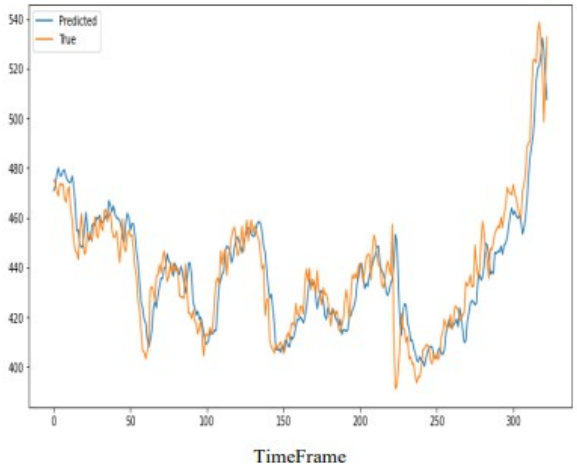

Fig 2. LSTM Prediction Shows the similarities in the predicted and actual values of the normalized stock price using LSTM approach. Plot with matplotlib.

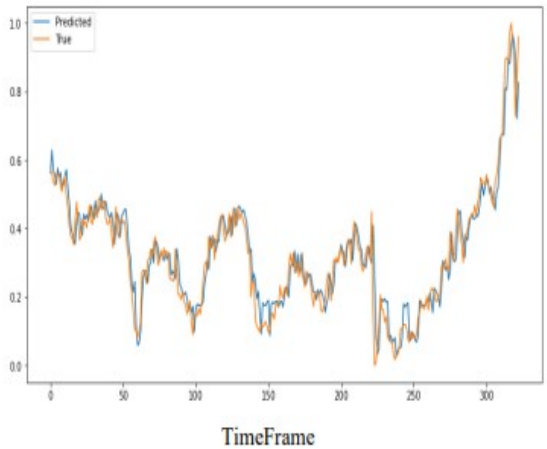

Fig 3. XGBoost Prediction Shows the similarities in the predicted and actual values of the normalized

stock price using XGBoost approach. Plot with matplotlib.

V. CONCLUSION

Stock Clue is a platform that has a variety of machine learning models, including text-based and series-based algorithms, that are used to anticipate stock price fluctuations. The accuracy of stock prediction models based on machine learning models is almost adequate for real-world use in most cases. In order for customers to make informed investment decisions, the news vector data must offer an indicator of stock volatility. XGBoost offers more accurate findings since it incorporates both recent news and time series information. The accuracy of stock forecasting using machine learning is now 90 percent, according to the latest data. Small-scale traders who do not have the time to learn about stocks may take use of this instrument at a low cost since the majority of the systems are open source.

REFERENCE

[1] Lei Shi, Senior Member, IEEE, Zhiyang Teng, Le Wang, Yue Zhang, and Alexander Binder – "DeepClue: Visual Interpretation of Text-based Deep Stock Prediction- ". IEEE Transactions On Knowledge And Data Engineering

[2] Yangtuo Peng and Hui Jiang – "Leverage Financial News to Predict Stock Price Movements Using Word Embeddings and Deep Neural Networks ". Keele Street, Toronto, Ontario, M3J 1P3, Canada

[3] WasiatKhan ,Mustansar Ali Ghazanfar , Muhammad Awais Azam , Amin Karami , Khaled H. Alyoubi and Ahmed S. Alfakeeh – "Stock market prediction using machine learning classifiers and social media, news ". Journal of Ambient Intelligence and Humanized Computing

[4] Xianghui Yuan JinYuan ,Tianzhao Jiang , and Qurat Ul Ain. "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market". IEEE Access.

[5] Jimgyi Shen and M OmairShafq, Shen and Shafiq J. "Short- term stock market price trend prediction using a comprehensive deep learning system". Big Data

[6] Guangyu Ding and Liangxi Qin. "Study on the prediction of stock price based on the associated network model of LSTM". International Journal of Machine Learning and Cybernetics (2020)

[7] Abdalraouf Hassan (Member, IEEE) AND Ausif Mahmood (Senior Member, IEEE) – "Convolutional Recurrent Deep Learning Model for Sentence Classification ". March 28, 2018.

[8] Yash Sharma, Gaurav Agrawal, Pooja Jain, Member, IEEE,Tapan Kumar Senior Member IEEE Indian Institute of Information Technology, "Vector Representation of Words for Sentiment Analysis Using GloVe". Kota-2017 International Conference on Intelligent Communication and Computational Techniques (ICCT) Manipal University Jaipur, Dec 22-23, 2017

[9] ARgukt, Nuzvid, , DR. T. V. Rajanikanth, Pantangi Rajashekar. " Sentiment analysis using harn algorithm-D V Nagarjana Devi". International Journal 2020.