StockClue: Stock Prediction using Machine Learning

A M Pranav, Sujooda S, Jerin Babu,
Amal Chandran
Department of Computer Science and Engineering
College of Engineering Perumon

Anoop S
Assistant Professor
Department of Computer Science and Engineering
College of Engineering Perumon

Abstract— **The stock market, sometimes known as the stock exchange, is one of the most complex and sophisticated ways to do business. The stock market is volatile, but it is also one of the most effective methods to make big profits when approached with discipline. Small businesses, brokerage firms, and financial sectors all rely on this entity to generate revenue and distribute risks. This work intends to use open-source libraries and pre-existing methods to create machine learning models in a WebApp to forecast future stock prices for exchange, in order to help make this volatile kind of commerce a little more predictable. To avoid the conventional method and avoid getting the outcome completely based on numbers, a system to incorporate text-based machine learning model and pattern recognition is implemented. The objective is to create a platform for small and amateur traders, from the existing stock prediction models considering variables like news articles, stock volume, previous close, etc. to predict the future stock market values.**

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

I.INTRODUCTION

Stock Market is one of the oldest methods where a normal person would trade stocks, make investments and earn some money out of companies that sell a part of themselves on this platform. Stock market provides a platform for almost all major economic transactions in the world at a dynamic rate called the stock value which is based on market equilibrium. This system proves to be a potential investment scheme if done wisely.

Predicting this stock value offers enormous profit opportunities which are a huge motivation for research in this area. Even a fraction of a second's knowledge of a stock's worth can result in large earnings. Similarly, in the repeated context, a probabilistically correct prediction might be highly profitable. This attractiveness of finding a solution has prompted researchers, in both industry and academics to find a way past the problems like volatility, seasonality and dependence on time, economics and rest of the market. However, the platform's prices and liquidity are highly unpredictable, which is where technology comes in to aid.

In the past, one of the more active study areas was stock market prediction. Small companies, brokerage firms, and banking sectors all rely on the stock market to generate money and so spread risk. As a result, predicting the stock value is a top priority. If we use crowd computing to graph the stock exchange price, we will receive an approximate answer to a real-life graph, but it will be an extremely slow process. Trading algorithms can now more accurately predict stock price movements thanks to recent advances in deep learning. Unfortunately, there is a large gap in the implementation of this breakthrough in the real world. Moreover, these advanced technologies are rarely used to benefit small-scale dealers.

The objective behind "StockClue" is to create an interactive web-app with effective ML models that could help many small-scale investors to invest for a long run. The app will be a ready-made site with all stock details like candle-stick graph, open price, close price, volume, relevant news, etc.

II.RELATED WORKS

The design of the neural network model used by Lei Shei et al. [1] based on news data and twitter tweets. A list of keywords for each organization is kept to match each piece of news to the appropriate stocks (e.g., Apple: AAPL, AAPL.O, APPLE, AAPL.N, Apple Inc, etc.). By matching the firm's hashtags in the tweet content, the stock-related tweets were extracted using the Twitter API. The model's purpose is to forecast a stock price y that is close to the firm's actual stock price y. DyNet v1.0, a neural network software library specialised for natural language processing applications, is used to construct DeepClue.They considered S&P 500 stocks in the US stock market from 2006 *to 2015*. Their historical prices are acquired from Yahoo Finance and financial news from Reuters and Bloomberg.

Yangtuo Peng et al. [2] developed a model that computes the closing prices of the preceding five days to create an input feature vector for DNNs. The model searches all financial publications for sentences that reference at least one stock name or public firm. Each sentence is sorted into samples and labelled with the original article's publication date and the relevant stock name. Each example includes a list of sentences published on the same day and mentioning the same stock or firm. Each sample is also labelled as positive ("price-up") or negative ("price-down") depending on the closing price the next day. The DNN is applied, which has hidden layers (each with 1024 hidden nodes). The historical price feature is used as a baseline, and additional financial news-derived elements are put on top of it. The DNN outputs are categorised by the dates of the samples in

the test set. All unseen stocks' predictions are compared to the actual stock movement the next day. The financial news data used in this paper are from Reuters and Bloomberg. The historical stock security data comes from the CRSP database (Center for Research in Security Prices).

Wasiat Khan et al. [3] have selected financial news Business Insider for analysis using Stanford sentimental analysis package of Stanford NLP giving positive or negative points for positive or negative words.

Xianghui Yuan et al. [4] proposed a model to forecast the stock's excess returns for the following month. The financial report, daily opening prices, closing prices, volumes, and other data of the A-share market over an eight-year period are used to acquire 60 attributes to be utilized as input to the model.

Jingyi Shen et al. [5] created a model to find the price trend by comparing the current closing price to the closing price of n trading days ago when labelling data. They use LSTM for time-series prediction that ensures the prediction model can capture both complicated hidden patterns and time-series-related patterns. This dataset consists of 3558 stocks from the Chinese stock market also data collected through the open-sourced API and leveraged a web-scraping technique to collect data from Sina Finance web pages, SWS Research website.

Guangyu Ding et al. [6] proposed an associated deep recurrent neural network model with multiple inputs and multiple outputs based on long short-term memory network. The associated network model can predict the opening price, the lowest price and the highest price of a stock simultaneously. The experimental data in this paper are the actual historical data downloaded from the Internet, Shanghai composite index 000001 and the others are two stocks of PetroChina (stock code 601857) on Shanghai stock exchange and ZTE (stock code 000063) on Shenzhen stock exchange.

Abdalraouf Hassan et al. [7] the proposed framework combines joint CNN and RNN framework with a set of feature maps learned by a convolutional layer with long-term dependencies learned via long-short-term memory and use an unsupervised neural language model to train initial word embeddings tuned by deep learning network then the pre-trained parameters of the network are used to initialize the model. The performance of the proposed model was evaluated on the Stanford Large Movie Review dataset (IMDB) and the Stanford Sentiment Treebank dataset (SSTb) derived from Rotten Tomatoes movie reviews.

Yash Sharma et al. [8] implemented Glove and gives the possible use of it in sentiment analysis. The word vectors obtained from Glove method is fed into RNN and sentiment analysis is doing binary classification (Positive and Negative Sentiments).

D V Nagarjana Devi et al. [9] proposed HARN algorithm, an unsupervised learning method which uses basic structure of the sentences, domain dictionaries and pre-defined polarities to classify the given sentence.

III.PROBLEM STATEMENT

The accuracy of the existing stock market prediction models is relatively low because only a small dataset is used for training, the results will be less accurate. There is still a need to continually explore more new features that are more predictable. Even though multiple algorithms exist, there is no real-life implementation of these ideas for the beneficial of people. Efficient algorithms should be made available with easy accessibility and interface.

IV.PROPOSED SYSTEM

The proposed method involves determining an interactive online platform (Web App) for stock traders to use in order to forecast future stock market values. The Web App also shows market prices, volume, and associated statistics, as well as the selected stock's prediction. The goal is to create a platform with many efficient stock market machine learning methods. Individually learned stock prediction parameters include Finance News, and Stock data.

The entire Web App can be divided into the following sections:

1. Login: WebApp has a page for user authentication where people can login to the app to view the live status and prediction of various stocks. The user can also sign up if login for the first time.

2. Live Page: The live page has a slider with listed stocks, a live stock price display, and predicted values. Users can select the wanted stocks from the slider and the corresponding live stock values, predicted graphs, and ML model results will be displayed.

3. News List: The current news about the particular stock will be displayed in the new bar which the user can view and also navigate to the news article by clicking the link.

The Current news is fetched from the cloud and displayed in the web app, the same news is also vectorized to its positivity or negativity factor and displayed to analise its impact on the stock price. The closing price of the market is collected to predict the future prices using LSTM. It is then graphically represented in the web app. The XGBoost prediction model collects the OHLC (open, high, low, close) price along with volume and news vector to predict the following days trends. It is also illustrated in the web app through graph.
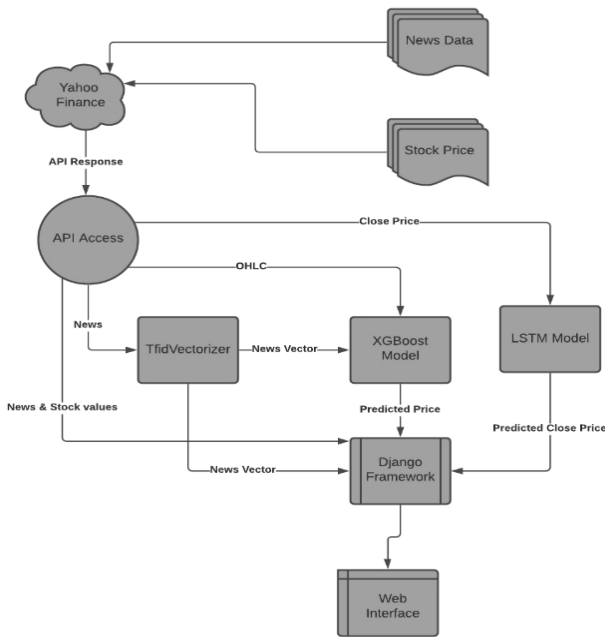
Fig 1 Stock Clue Architecture Design
Complete backend design of the web app

V.IMPLEMENTATION

The backend of the entire project is implemented using python and Django framework. PostgreSQL manages the user data and login information. HTML, CSS and JavaScript is used effectively to make a reactive web app. Chart.js library is used to display the stock charts.

Live Data: The ML models need current data like news, volume, etc. The web app also needs the data to display the current stock graph. Thus, the data is fetched using Yahoo Finance API and nsepython package

ML Models: The app has 3 stock prediction models. Sentiment Vector: Finds the trend of the current news of the stock. LSTM finds the closing price of next 30 days. XGBoost predicts the stock price with news data, volume, previous close, open, etc.

1)LSTM: The LSTM model is used to predict the stock closing price for the next 30 days from the current date. The Long-short term memory model is trainned with the closing price the particular stock to predict the future closing patters.

The 20 years of closing price of 'Infosys' was taken for LSTM. 80% of the data was trainned and 20% was used for testing. The model showed an accuracy of 99% accuracy with the test data. The model was then saved with the webapp. The dataset was downloaded from kaggle.

2)News Vector: The TfidfVectorizer from scikit-learn is used to analyze the current news data. The model takes the news data as input and gives a vectorized value of the news. It gives the positivity and negativity of the news. It is then incorporated with the webapp. The current news data is fetched through yahoo finance API.

3)XGBoost: The XGBoost algorithm gives the most accurate and reasonable prediction. The model takes news vector, open price, close price, volume of 3 days as input and predicts the closing price for the next 3 days.

The model was trained with 5 years of data. The news data is extracted by web scraping and the quant data is from Kaggle. The dataset was pre-processed to combine the news data with the quant on the basis of date. 80% of the data was used for training and 20% was used for testing. The model showed 90% accuracy. It is then saved with the web app.

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.
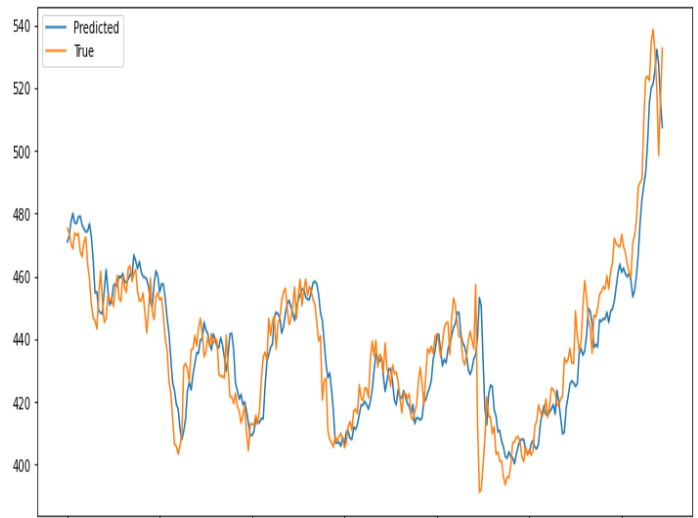
TimeFrame

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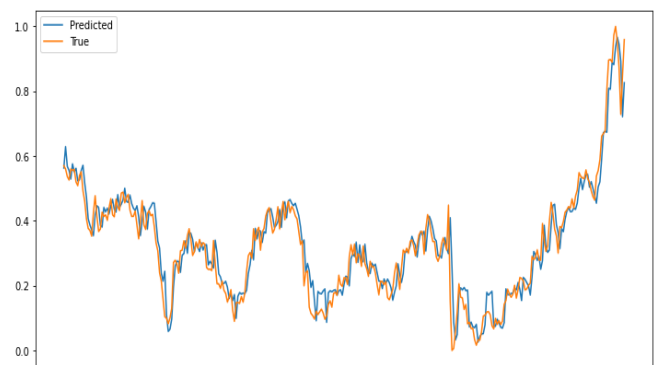
TimeFrame

Fig 3 XGBoost Prediction
Shows the similarities in the predicted and actual values of the normalized stock price using XGBoost approach. Plot with matplotlib.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCREIS - 2021 Conference Proceedings

Fig 4 Sentimental Vector
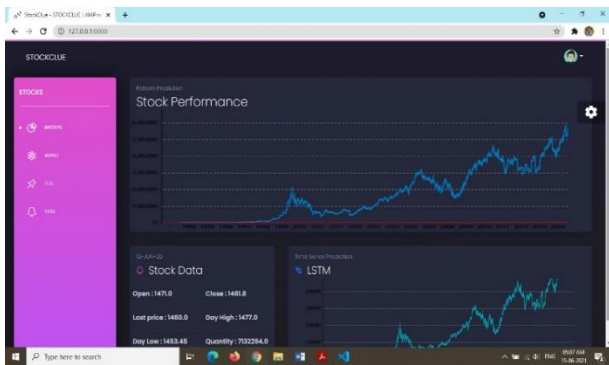The news data vectorized using TfidfVectorizer

Fig 5 StockClue Webapp
The Django interface showing all the stock predictions.

Models	Technical Analysis Accuracy	Fundamental Analysis Accuracy
StockClue	99%	90% - 95%
Stock market prediction using machine learning classifiers and social media, news [3]	N/A	75%-80%
Deepclue: visual interpretation of text- based deep stock prediction [1]	N/A	90%-95%
Leverage financial news to predict stock price movements using word embeddings and deep neural networks [2]	N/A	63%
Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms [4]	57%	N/A
Short-term stock market price trend prediction using a comprehensive deep learning system[5]	89%	N/A

Table 1 Comparison Table
Comparison of our model with other models stated in other
approaches from various journals.

VI. CONCLUSION

Stock Clue is a platform that has many ML models including text based and series based that helps to forecast stock price variations. Stock prediction accuracy using machine learning models are pretty much sufficient for a real-life implementation. The news vector data provides an indication of stock volatility, allowing users to invest effectively. Because the XGBoost model mixes news and time series data, it produces more accurate results. Stock prediction accuracy using machine learning is currently at 90 percent. Small-scale traders do not have must time in learning a stock can use this tool with cheap price as most of the system are open source.

REFERENCES

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

[2] Leverage Financial News to Predict Stock Price Movements Using Word Embeddings and Deep Neural Networks-Yangtuo Peng and Hui Jiang - Department of Electrical Engineering and Computer Science York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3, Canada

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

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

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

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

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

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

[9] Sentiment analysis using harn algorithm-D V Nagarjana Devi - Assistant Professor, IIIT, Rgukt,Nuzvid , DR.T.V.Rajanikanth-Professor,Snist , Hyderabad.Tg , Pantangi Rajashekar - UG Student, IIIT,Rgukt,Nuzvid Gangavarapu Akhil - UG Student , IIIT, Rgukt,Nuzvid