Stock Prediction and Recommendation with Deep Learning Analysis

Jiaqing Chen [1] **Wannuo Sun** [1]

Abstract

Stock market prediction is always an attractive and challenging task among investors as it contributes to developing effective strategies in stock trading. This project developed a comprehensive stock prediction system employing Long Short-Term Memory (LSTM), Temporal Convolutional Networks (TCN), and several traditional machine learning models to forecast long-term stock trends and short-term up/down phenomena. Our model takes the publicly available historical stock price as long as topic-related tweets to drive predictions by both history and public knowledge. Empirical experiments show that our models achieve around 80% validation accuracy. We built a dashboard application to offer our prediction results and serve for public usage.

1. Introduction

For many individuals and families, making investments is one of the most efficient methods to take advantage of a growing economy and stay ahead of inflation. With access to massive information nowadays, people can investigate stock markets more insightfully than before. Big data and advanced artificial intelligence tools will be beneficial for people to make decisions and choose investment strategies. As a result, the goal of our system is to take advantage of deep learning tools to provide a comprehensive stock analysis and predictions and, correspondingly, recommendations for investors.

Different factors and models of stock prediction have been extensively studied by researchers in both academia and industry. Some researchers rely purely on historical price and volumes data to make history-driven predictions, while others develop tools to combine it with financial factors like tweets, news and make use of knowledge graphs to make knowledge-driven predictions. With the development of deep learning, more advanced models are incorporated into market prediction. For example, Long short-term memory (LSTM) model (Chen et al., 2015; Nelson et al., 2017) is extensively used in time-series analysis and is particularly suitable for the stock price. Recently, a cutting-edge model, Temporal Convolutional Networks (TCN), combines the advantage of Convolutional Neural Network (CNN) and recurrent neural network (RNN), leading to a new generation of time-series prediction (Bai et al., 2018). Predicting the future market becomes more possible.

Our system incorporates both history and public knowledge to drive stock analysis from the perspective of long-term and short-term. For input space, we involved tweets in providing public sentiment with the assumption that behavior economics affects the market moving. For modeling tools, we leveraged several machine learning and deep learning models to assist in prediction, including hot LSTM and cutting-edge TCN models. In general, our work includes (1) building an LSTM model on the historical stock price for a long-term trend prediction, (2) self-designing a deep ensemble consisting of TCN, LSTM and several traditional machine learning models and training on tweets sentiment as long as stock vectors to offer short-term up/down judgment, (3) developed a dashboard application to demonstrate our forecasting and serve for investor's usage.

The remaining of the report is organized as follows. Section 2 discusses the related work in stock market prediction. Section 3 presents our proposed system and detailed methodologies. Section 4 introduces our dataset usage. Section 5 empirically demonstrates the model's effectiveness through validation testings. Section 6 shows our dashboard application and main components. Section 7 discusses the critical parts and envisions possible future works. Section 8 summarizes our work as a whole.

2. Related work

This section introduces the related work from the stock market prediction method using technical analysis and fundamental analysis.

2.1. Technical Analysis

Some work makes time series stock prediction based purely on history price and tends. Given the history price and

[1]Department of Computer Science at Columbia University, New York, NY, United States. Correspondence to: Jiaqing Chen, Wannuo Sun <jc5657@columbia.edu, ws2591@columbia.edu>.

trends, we can generate statistical indicators along with the original data as our input for different models, including basic machine learning models and advanced deep learning models. (Armano et al., 2005) introduces a hybrid genetic-neural architecture for stock index forecasting. (Abraham et al., 2001) makes use of a neural network for one day ahead stock forecasting and a neuro-fuzzy system for analyzing the trend of the predicted stock values. (Huang & Tsai, 2009) hybridizes Support Vector Regression (SVR) with the self-organizing feature map (SOFM) technique to predict the stock market price index. (Ariyo et al., 2014) uses the ARIMA model to build a stock price predictive model with an extensive process. (Zhou et al., 2018)we use Long Short-Term Memory (LSTM) and convolutional neural network (CNN) for adversarial training to forecast the high-frequency stock market.

2.2. Fundamental Analysis

However, purely technical analysis shows drawbacks when dealing with abrupt changes in the stock market. Fundamental analysis, which combines historical data with financial factors like tweets and news, has been developed to address this issue. Many event-driven methods utilized the events extracted from news, social media, and discussion board to forecast the stock trend in recent years.

(Ding et al., 2015) represents events extracted from news text as dense vectors and uses a deep convolutional neural network to model both short-term and long-term influences of events on stock price movements. (Deng et al., 2019) uses the TCN model, and combines event embeddings, which are obtained by extracting structured events from financial news and utilizing external knowledge from the knowledge graph, and price values together to forecast stock trends.

Based on (Bollen et al., 2011), which shows the results that the accuracy of DJIA predictions can be significantly improved by the inclusion of specific public mood dimensions but not others, further works also use sentiment in news or texts in predictions. (Liu et al., 2019) proposes a new method to calculate the correlation by using the enterprise knowledge graph embeddings and employs Gated Recurrent Unit (GRU) model to combine the correlated stocks' news sentiment, the focal stock's news sentiment, and the focal stock's quantitative features to make predictions. (Xu et al., 2021) proposes a relational event-driven stock trend forecasting (REST) framework. It models the stock context and learns the effect of event information on the stocks under different contexts, constructs a stock graph, and designs a new propagation layer to propagate the effect of event information from related stocks.

3. System Design

In this section, we first present the overview of our system for stock analysis and prediction. Then we introduce each component in detail.

3.1. System Overview

In order to provide a comprehensive analysis and prediction of stock performance, we focus our eye on both short-term and long-term perspectives, combining technical analysis and fundamental analysis.

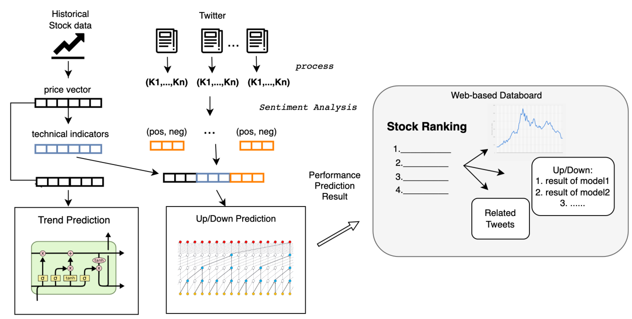


Figure 1. System Framework

In general, we divided our analysis into two parts, as in Figure 1. First, we built an LSTM model for a 15-day window long-term trend prediction. The model is trained with 5-year stock price vectors and demonstrated great performance in describing the stock trend. Second, we designed an ensemble model consisting of several machine learning tools and deep learning models for a next-day up/down judgment. We extracted technical indicators from the stock vector, collected stock-related tweets, and performed sentiment analysis to conclude the public mood. We built the training set by concatenating the stock price, typical technical indicators, and sentiment results. In this way, we enable our prediction to be driven by both history and public knowledge. Finally, we provided our short-term and long-term combined forecasting on a well-organized dashboard.

3.2. Long-term View: History-driven Trend Prediction

History-driven trend prediction focuses on historical stock open prices to predict the future trend. Here, we built an LSTM model with three layers and dropout techniques. With the assumption that an arbitrary rise or down is the event with low probability in the stock market, the stock price is correlated with the historical price, and such a correlation fades with time interval increases. Inspired by this fact, we focus on the previous 30-day price open price in history to predict the 31^{th} day open price. In order to have a full 15-day forecasting, we adopted a roll-over prediction which means the new prediction result will be involved in the next-day prediction, as depicted in Figure 2. Under this methodology, the 15^{th} day price will be predicted on an actual half price and a half prediction results in the previous day. It could be demonstrated that even though the predicted price would not be so accurate, the moving trend is generally matched.

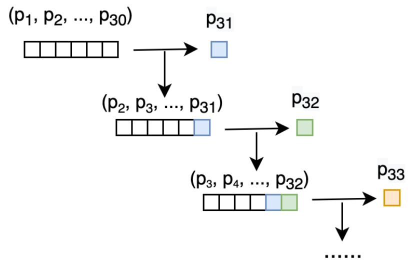


Figure 2. Roll-over Prediction

3.3. Short-term View: Knowledge-driven Next-day Up/Down Prediction

This model aims to forecast whether the target stock index would go up or fall down in the next-day. We constructed a binary prediction, where 1 denotes rise and 0 denotes drop, defined as

$$y_i = \begin{cases} 1, & p_{i+1} \leq p_i \times (1+\alpha) \\ 0, & p_{i+1} > p_i \times (1+\alpha) \end{cases}$$

Here α remains for further flexible changes. For example, if we would only invest in the stock with higher than 5% rise, α could be assigned with 0.05 to satisfy needs. Without loss of generality, we define $\alpha = 0$ here. We involve both technical analysis and fundamental analysis here to drive the binary prediction.

3.3.1. Stock Price and Moving Structuralization

It is still necessary to deploy the historical stock price following the technical analysis principle. Additionally, we encoded the history moving trend into vectors here via technical indicators.

Stock price vector. Besides open price, we included a full picture of the original stock price vector, consisting of open price, close price, high price, low price, adjusted low price, and the trading volume at the i^{th} day. We expect the differences in these values offer a hint for the next-day trend.

Technical indicators. Technical indicators are heuristic or pattern-based signals produced by attributes of stock, like price and volume, for traders to perform technical analysis. For example, the Average Direction Index (ADX) helps the trader to determine the strength of a trend and indicate whether a trade should be buying or selling; Stochastics (STOCH) offers traders a different approach to calculating price oscillations by tracking how far the current price is from the lowest low of the last X number of periods. We selected and calculated ten commonly used technical indicators through `finTA`. The calculation is performed on the whole history data such that the history moving is embedded in the indicators for each daily vector.

3.3.2. Tweet Sentiment Extraction

Inspired by the premise of behavioral economics that the public sentiment strongly correlated with their behavior and further the market sentiment(Mittal & Goel, 2012), we analyzed the public mood through Twitter and encoded the sentiment to our input space.

Tweet Crawler. The old tweets for each day in the past five years was crawled based on keyword search. We further process the tweets via

- Emoji and non-word content removing with `tweet-preprocessor` package.
- Keeping only English employing `langdetect` package.
- Text clean, stemming and lemmatization with Natual Language Processing (NLP) tools.

Sentiment Analysis. We applied Valence Aware Dictionary for Sentiment Reasoning (VADAR) model to justify the positive level and negative levels of a tweet. VADAR is a lexicon and rule-based feeling analysis instrument that is explicitly sensitive to suppositions communicated in web-based media(Hutto & Gilbert, 2014). It is popular in extracting sentiment in text-based data. We assume that a higher positive proportion may indicate a higher possibility of stock going up and vice versa. For example, Figure 3 is the sentiment analysis result of AAPL and GOOG on a specific day. It is clear that more positive content appears in AAPL while negative take up the all in GOOG. Intuitively, we could predict that GOOG would go down the next day with a high possibility.

In particular, we calculated the text ”Compound score”. We define a tweet has a positive mood if its compound score is no less than 0.05, and a tweet is negative if the score is no larger than -0.05. Suppose there are N_p positive tweets and N_q negative tweets, we encoded the proportion factor simply as

$$Proportion_p = \frac{N_p}{N_p + N_q}$$

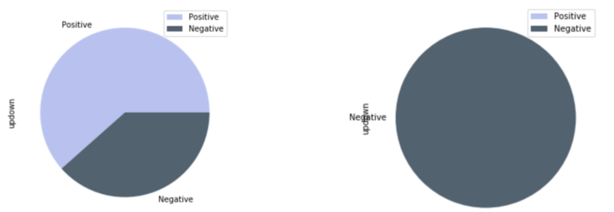


Figure 3. Sample of Sentiment Analysis result of AAPL (left) and GOOG (right)

$$Proportion_n = \frac{N_q}{N_p + N_q}$$

3.3.3. Knowledge-driven Multi-source Concatenation

Based on above extraction information, the input vector is organized as

$$P_i = [P_{p_i}, P_{t_i}, P_{s_i}]$$

where we have P_{p_i} represents the price vector for day i, P_{t_i} denotes the technical indicators, and P_{s_i} is the public sentiment obtained from tweets at day i. We performed Min-max normalization to guarantee feature balance.

3.3.4. Model: Deep Ensembles

Instead of building a single deep learning model, we designed a self-made ensemble model that integrates the cutting-edge time-series deep model, Temporal Convolutional Network (TCN) and Long short-term memory (LSTM) model, and a series of traditional machine learning models. The stock market is usually hard to forecast due to several random moving and arbitrary changes. We expect models built with different logic could capture various features of a stock period price.

Temporal Convolutional Network (TCN). Benefited by the feature of no future information "leakage" and the ability to keep a long effective memory, TCN has great potential in analyzing time-series stock data. It has been demonstrated to be better than LSTM and GRU on a vast range of tasks like MNIST. Also, due to its parallelism, the model could be built effectively(Bai et al., 2018; Lea et al., 2016; 2017; Deng et al., 2019). In this project, we adapt the generic TCN architecture described from (Bai et al., 2018) as the starting point. As depicted in Figure 4, TCN is based on a 1D fully convolutional network. It accepts input sequences with any length and provides an output of the same size. Zero-padding is added to keep all hidden layers the same size as the input layer. For stock binary forecasting, a dense layer with nodes in a number of classes is attached to the output of the TCN. Several techniques employed in TCN, like Dilated Causal Convolutions and Residual Blocks (Deng et al., 2019) distinguishes the TCN from other time-series models.

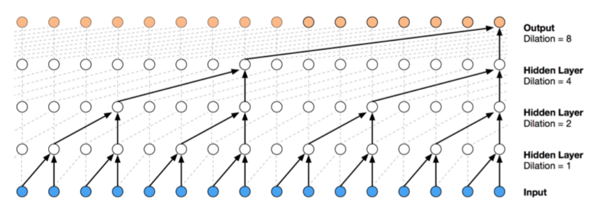


Figure 4. Sample of TCN model from (Bai et al., 2018)

We implemented TCN models with open-source implementation and fine-tuned it based on our specific needs. Here, we leveraged early stop to prevent overfitting during the training process. As rising and dropping are both normal cases in the stock market, we gained a relatively balanced classification dataset, such that we mainly applied validation accuracy to monitor model behavior in both early stopping choice and further model evaluation.

Long short-term memory (LSTM) LSTM is a hot model in recent years which is developed based on a recurrent neural network. It could process not only single data points but also entire sequences of data such that it demonstrated great potential in time-series problems, especially in stock price (Nelson et al., 2017; Chen et al., 2015). We employed it in both long-term prediction and short-term prediction here. We implemented it with `Keras` and built three layers to draw the prediction.

Traditional Machine Learning model. Several classical machine learning models have their personal benefits in capturing stock features. We built K-nearest Neighbor (KNN), Random Forest, Support Vector Machine (SVM), Logistic Regression, and Gaussian-based Naive Bayes to serve in our model pool. All of them are implemented with `sklearn` package.

Voting Decision. We employed majority voting to determine the final prediction results. In other words, we collected all the prediction results and took the majority as our final output. In particular, we assign voting participants to only models with a validation accuracy larger than 0.75. In this way, we guarantee that the decision is not coming from a single model but several "dependable" models.

4. Dataset

Our dataset included publicly available historical stock prices and self-crawled tweets.

4.1. Time-series Price Dataset

The training price dataset contains daily value records of each company stock and S&P500 index, with a timespan from 01/01/2018 to 05/06/2022. Stock price data are from Yahoo! Finance and extracted using `yfinance` API. Note that they are cleaned for bank holidays and aligned with tweets by time. Examples are shown in figure 5.

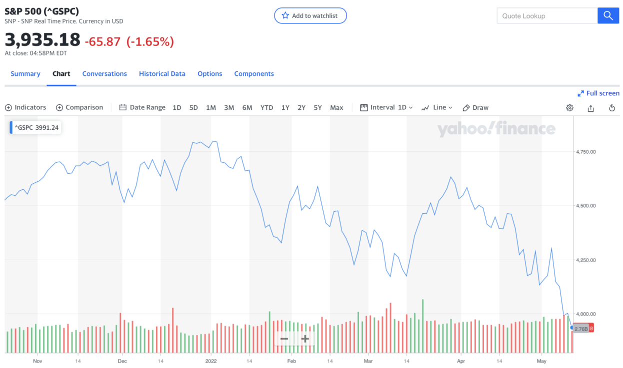


Figure 5. Example of time-series stock price

4.2. Topic Related Tweets

The tweet dataset is composed of historical tweets from Twitter with the timespan from 01/01/2018 to 05/06/2022. For each stock trading day, the top 60 tweets are extracted using `sntwitter.TwitterSearchScraper` API. Example are shown in figure 6.

date	tweet_0	tweet_1	tweet_2	tweet_3	tweet_4	tweet_5	tweet_6	tweet_7
2018-1-1	Number of leading stocks today: 0 SPYQQQ $F...	There is a 40% chance Apple will acquire Netfl...	Tuesday: $WATT$,TSLA, $NVDA$,AAPL, $BABA	Here's your gift for the Holidays! Heat up you...	I just upgraded from an iPhone 6 to an 8 Plus....	Today is #NewYearsDay that means we all need t...	Is anyone else sick of seeing $AAPL pay prompt...	My favorite day of the year, going through all...
2018-1-2	If you type “Sell” in Google Finance, it'll br...	Citigroup comments that there was a "40% chanc...	"Nay, we don't need patents...or a strong pate...	$180 for$AAPL? This month it's possible more ...	Some of my *best performing #STOCKS TODAY 1/2:...	Up Moves:\n+3.7% XME\nFXI\n+3% $+3.5$ $EWZ\n+...	$SIEB$YY $WUBA$X $NFLX$FSLR $AAPL$FB $ERX ...	Number of leading stocks today: 74 SPYQQQ $...
2018-1-3	U can make ur iPhone 6 basically new again by ...	$LITE best way to play surging 3D sensing mark...	The Nasdaq Composite closed above 7,000 for th...	$BABA$ TO200-250MU 90AAPL 220NVDA $3...	Long in BACAAPL CSXJPM KHCKO #WallStreet	In long on BACAAPL KHCKO #WallStreet	Number of leading stocks today: 76 SPYQQQ $...	Learning to trade profitably and comfortably i...
2018-1-4	$AAPL Says all IOS devices affected. All MAC ...	$AAPL Headlines hitting known meltdown issues ...	#AAPL is undervalued. Senseless debates genera...	If AAPL doesn't hit a $1Trillion MKTcap, I wil...	AAPL store holiday sales broke some kind of re...	$AAPL news addd 6 cents to QQQ post market #Lucky	Guys, do you also think $AAPL should buy$NFLX...	Over half a billion users visit the iOS App St...

Figure 6. Example of tweets collected

5. Experiment

In this section, we will demonstrate the experiment results of models in two parts.

5.1. Environment

The project is implemented on a computer system consisting of an Intel Core i7-9700 CPU running at 2.8 GHz, 16 GB RAM, and a computer system consisting of an Intel Core i7-9700 CPU running at 2.6 GHz, 32 GB RAM. All implementations, including data collection, pre-processing, model building, and dashboard development, are completed in Python. Some open-source library usage and implementation details of each task are specified in the corresponding sections.

5.2. Experiment on long-term trend prediction

Long-term trend prediction is based on historical open prices and built with a three-layer LSTM model. In general, although the predicted price could not be as accurate as of the actual price, the general trend is matched. Some validation examples are shown in Figure 7.

Figure 7. Experiment result of long-term trend prediction

5.3. Experiment on short-term up/down prediction

Experiments demonstrate that, with fully-informed input, models in the deep ensembles have great performance in predicting next-day up/down of stock price. Models for different stocks show various prediction power dependent on specific stock's predictability. However, on average, we achieved around 80% validation accuracy and more than 65% for the hardest stock. Detailed model comparisons are listed in Table 1. Among all models, TCN shows the greatest performance, and Random Forest also suits. In general, due to the balance of the dataset, models perform well in both predicting ”up” and predicting ”down”. Example heat-maps of confusion matrix of some models as Figure 8 shows it.

6. Dashboard Application

A web-based dashboard is built to demonstrate our results and offer recommendations for investors. Two main pages are offered in the dashboard as shown in Figure 9:

- Overview Page: Introduce the dashboard and provides

Table 1. Short-term up/down model (average) performance

MODEL	ACCURACY	F-1 SCORE
TCN	**86.36**	**82.06**
LSTM	81.82	77.82
KNN	77.84	76.23
SVM	83.52	81.79
Random Forest	**84.66**	**82.68**
LOGISTIC REGRESSION	81.23	77.64
NAIVE BAYES	76.13	73.52

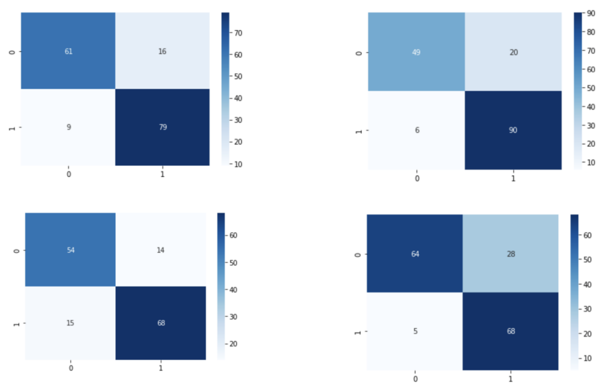


Figure 8. Experimental heatmap of short-term up/down prediction models

an overall analysis of several stocks.

- Company Stock Detail page: Lists all analyses for a specific stock, including its history stock performance, predicted trend, and predicted Up/Down class under each model.

Currently, we provide predictions on twelve technology companies: Google (GOOG), Meta (FB), Amazon (AMZN), Apple (AAPL), Microsoft (MSFT), IBM (IBM), Dell (DELL), Intel (INTC), Tencent (TCEHY), Cisco (CSCO), Sony (SONY), HP (HPQ). We compare them with SP 500 (GSPC), which is the market index, to conclude its relative performance in the market. The dashboard will automatically gather information and perform predictions upon opening it. The main components of the dashboard are elaborated below:

Stock analysis and prediction overview. As depicted in Figure 10, the overview offers a full picture of stock prediction results. It also served as our recommendation. We order the stock based on their predicted trend in 15 days, it is measured by its level of rise or down compared with today, i.e.

$$p_{trend} = \frac{p_{15} - p_0}{p_0} \times 100$$

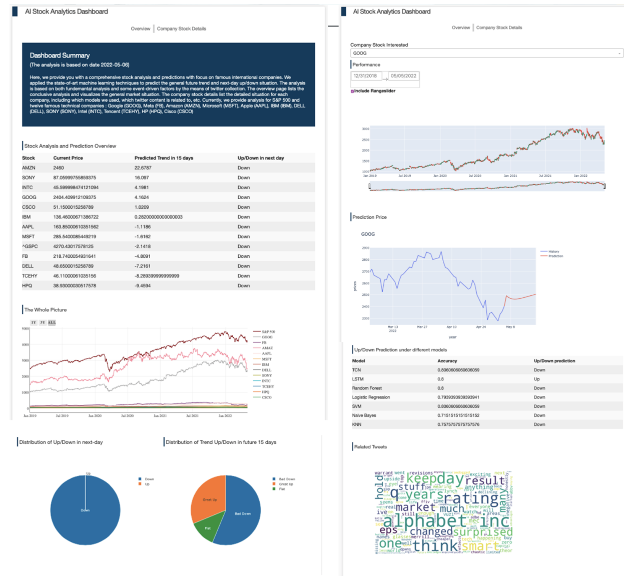
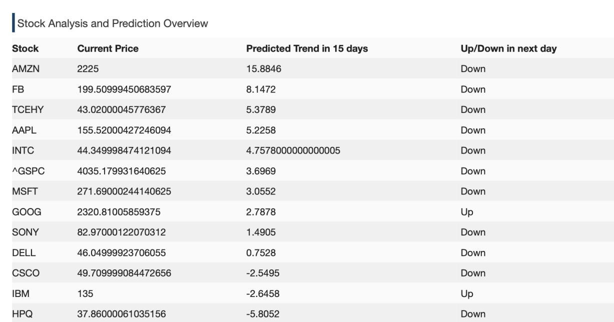


Figure 9. Dashboard Application

The results of the 15-day trend and the next-day up/down may be different. An "up" in the next day does not indicate a positive rise in the future 15 days, and a "down" does not indicate a negative down in future 15 days as well. We expect investors to focus on both parts to make their investment decisions. In general, we would recommend the stock with performance greater than SP 500, which is listed above GSPC.



Stock Analysis and Prediction Overview

Stock	Current Price	Predicted Trend in 15 days	Up/Down in next day
AMZN	2225	15.8846	Down
FB	199.50999450683597	8.1472	Down
TCEHY	43.02000045776367	5.3789	Down
AAPL	155.52000427246094	5.2258	Down
INTC	44.349998474121094	4.7578000000000005	Down
^GSPC	4035.179931640625	3.6969	Down
MSFT	271.69000244140625	3.0552	Down
GOOG	2320.81005859375	2.7878	Up
SONY	82.97000122070312	1.4905	Down
DELL	46.04999923706055	0.7528	Down
CSCO	49.709999084472656	-2.5495	Down
IBM	135	-2.6458	Up
HPQ	37.86000061035156	-5.8052	Down

Figure 10. Overview in Dashboard

Market distribution. We demonstrated the market prediction via two pie charts for both short-term prediction and long-term forecasting, as shown in Figure 11. We defined a stock with a rise greater than 1 as "great up", a stock with a down greater than 1 as "bad down", and others as "flat".

Trend and up/down prediction results. Company Stock Detail page lists all analysis for a specific stock, including its history stock performance, predicted trend, and predicted

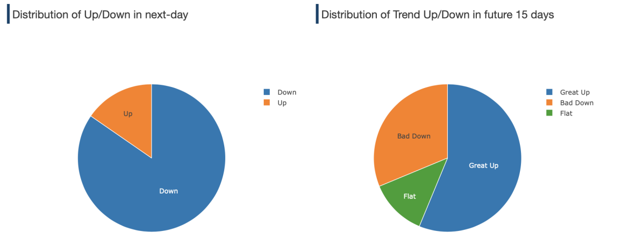


Figure 11. Pie chart analysis in Dashboard where the left one is based on next-day up/down prediction and the right one is based on long-term trend prediction

Up/Down class under every model, as shown in Figure 12 and Figure 13.

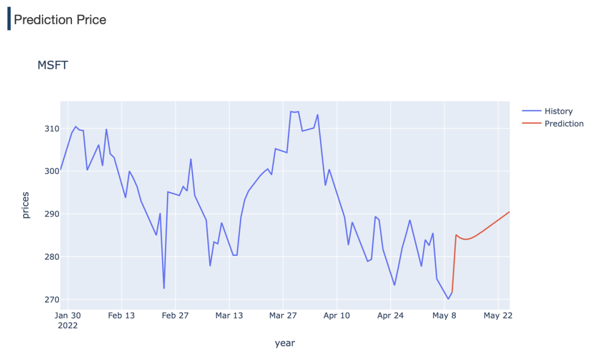


Figure 12. Long-term trend prediction in Dashboard

Up/Down Prediction under different models

Model	Accuracy	Up/Down prediction
TCN	0.8186046511627907	Up
LSTM	0.7906976744186046	Up
Random Forest	0.8325581395348837	Up
Logistic Regression	0.8325581395348837	Up
SVM	0.8232558139534883	Up
Naive Bayes	0.8046511627906977	Down
KNN	0.7813953488372093	Up

Figure 13. Next-day up/down prediction in Dashboard

Tweet wordcloud. We constructed a word cloud for each day's tweets information as depicted in Figure 14. We expect investors could quickly capture the public focus and topics on the specific stock.

Figure 14. Tweet wordcloud in Dashboard

7. Discussion and Future Work

With the objective of building customized models for each stock, the system scale is proportional to the stock amount. However, our proposed system is easily scalable, and all the models follow the same logic during predictions. Currently, we have models for twelve companies and the market index. As long as there is sufficient computing resource and storage space, the system can be enlarged to include analysis of hundreds or thousands of stocks. We hope our system could be served as a start to building a large, comprehensive, and fully-informed stock prediction system.

There are more future works that could be done as well. On the one hand, Twitter could be involved more to drive analysis. Currently, we only leveraged the sentiment part, but more information embedded in tweets remains to be explored. On the other hand, historical stock data and tweets are still not sufficient for stock prediction. Several factors have an impact on the stock market, ex. The correlation between stocks, the financial reports of individual companies, etc. If more elements could participate in the analysis, we could expect more accurate and reasonable forecasting.

8. Conclusion

Towards the objective of assisting investors in making an investment decision, we designed a system that incorporates both technical analysis and fundamental analysis to predict the stock market. We built an LSTM model on historical stock open price to predict the stock price for the next 15 days. We demonstrated that it indeed described the future trend. Furthermore, we involved tweet sentiment as long as the stock price vector and technical indicators to construct a short-term up/down forecasting. It achieves around 80% validation accuracy. Finally, integrating the analysis results, we built a dashboard application that demonstrates our results and provides stock recommendations for investors. With the feature of easy scalability, we hope our system could act as a start for a complete stock prediction system and serve as a powerful tool for investors to outlook the future stock market.

Acknowledgements

This work for EECS E6895 Advanced Big Data Analytics and AI course at Columbia University. We are gratitude for the professor's and teacher assistant's help and suggestions for our project.

References

Abraham, A., Nath, B., and Mahanti, P. K. Hybrid intelligent systems for stock market analysis. In *International Conference on Computational Science*, pp. 337–345. Springer, 2001.

Ariyo, A. A., Adewumi, A. O., and Ayo, C. K. Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp. 106–112, 2014. doi: 10.1109/UKSim.2014.67.

Armano, G., Marchesi, M., and Murru, A. A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, 170(1):3–33, 2005.

Bai, S., Kolter, J. Z., and Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Bollen, J., Mao, H., and Zeng, X. Twitter mood predicts the stock market. *Journal of computational science*, 2(1): 1–8, 2011.

Chen, K., Zhou, Y., and Dai, F. A lstm-based method for stock returns prediction: A case study of china stock market. In *2015 IEEE international conference on big data (big data)*, pp. 2823–2824. IEEE, 2015.

Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J. Z., and Chen, H. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Companion Proceedings of The 2019 World Wide Web Conference*, pp. 678–685, 2019.

Ding, X., Zhang, Y., , T., and Duan, J. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

Huang, C.-L. and Tsai, C.-Y. A hybrid sofm-svr with a filter-based feature selection for stock market forecasting. *Expert Systems with applications*, 36(2):1529–1539, 2009.

Hutto, C. and Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pp. 216–225, 2014.

Lea, C., Vidal, R., Reiter, A., and Hager, G. D. Temporal convolutional networks: A unified approach to action segmentation. In *European Conference on Computer Vision*, pp. 47–54. Springer, 2016.

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., and Hager, G. D. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–165, 2017.

Liu, J., Lu, Z., and Du, W. Combining enterprise knowledge graph and news sentiment analysis for stock price prediction. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

Mittal, A. and Goel, A. Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, 15:2352, 2012.

Nelson, D. M., Pereira, A. C., and De Oliveira, R. A. Stock market's price movement prediction with lstm neural networks. In *2017 International joint conference on neural networks (IJCNN)*, pp. 1419–1426. IEEE, 2017.

Xu, W., Liu, W., Xu, C., Bian, J., Yin, J., and Liu, T.-Y. Rest: Relational event-driven stock trend forecasting. In *Proceedings of the Web Conference 2021*, pp. 1–10, 2021.

Zhou, X., Pan, Z., Hu, G., Tang, S., and Zhao, C. Stock market prediction on high-frequency data using generative adversarial nets. *Mathematical Problems in Engineering*, 2018.