Stock Price Forecasting & Volatility Analysis

BTP Presentation

UNDER THE SUPERVISION OF

Mrs. Preeti Kaur

Nipun Jain, 2017UCO1543
Harshit Mahajan, 2017UCO1565
Mohit Motiani, 2017UCO1569

01. Introduction

02. Problems addressed

03. Brief discussion of related work

04. Approach
- Classical Models
- Using Sentiment Analysis

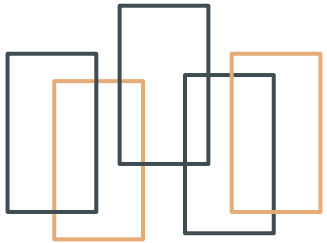
05. Results

06. Challenges, Learnings and Future Scope

07. List of Papers referred

08. Timeline

Introduction



Stock market prediction is an attempt to forecast the future trend of an individual stock, a particular sector or the market, or the market as a whole.

Approaches Considered-

- Classical Deep Learning Models
- Using Sentiment Analysis

Problem addressed

Classical Deep Learning based models

We try to contrast and compare the effectiveness of Classical Deep Learning based models to predict the stock prices for stocks of a specific company.Here we use the following models and their common variations in our thesis.

1. Vanilla RNN
2. Long Short Term Memory(LSTM) Model
3. Gated Recurrent Unit(GRU)
4. Convolutional Neural networks(CNN)

Volatility Prediction using Sentiment Analysis

According to EMH Theorem, Investors are not always rational while investing and are often dictated by Greed and Fear. Therefore, in order to capture the investor's mood, we are building a RNN(Recurrent neural network) which uses sentiment analysis along with historical stock price data to observe if it provides any improvement over the classical approach.

Brief Discussion of Related Work

Neural networks for stock price prediction (2018)

Ren-Jie Han, Yu-Long Zhou, Yue-Gang Song

Five neural network models, namely, back propagation (BP) neural network, radial base function (RBF) neural network, general regression neural network (GRNN), support vector machine regression (SVMR), least square support vector machine regression, are surveyed and compared with predictive capacity (LS-SVMR). They conclude that the BP neural network outperforms the other four models.

Stock Prices Prediction using Deep Learning Models (2019)

Jialin Liu, Fei Chao, Yu-Chen Lin, and Chih-Min Lin

Deals with the challenges of using deep learning models for predicting stock prices, since there is a lot of noise and confusion in stock price-related details. In order to denoise the data, they use (1-D) residual convolutional networks and in order to forecast the stock price, long-short term memory (LSTM) is then used.

Stock Trend Prediction using News Sentiment Analysis

Kalyani Joshi, Prof. Bharathi H.N., Prof. Jyothi Rao

Focuses on the famous theory of stock prediction i.e the Efficient Market Hypothesis. This project is about taking non-quantifiable data such as a company's financial news articles and forecasting the future market trend with the classification of news sentiment. They developed three distinct classification models. Observations show that in all forms of testing, RF and SVM perform well. Naive Bayes performs well, but not in contrast to the other two.

Sentiment Analysis of Twitter Data for Predicting Stock Market Movements

Venkata Sasank Pagolu, 2016

Highlights that social media nowadays are a true reflection of public perception and opinion on current affairs. Forecasts based on public opinions shared on Twitter. Two separate textual representations, Word2vec and N-gram, have been used in the present paper to examine public feelings in tweets.

Classical Approach

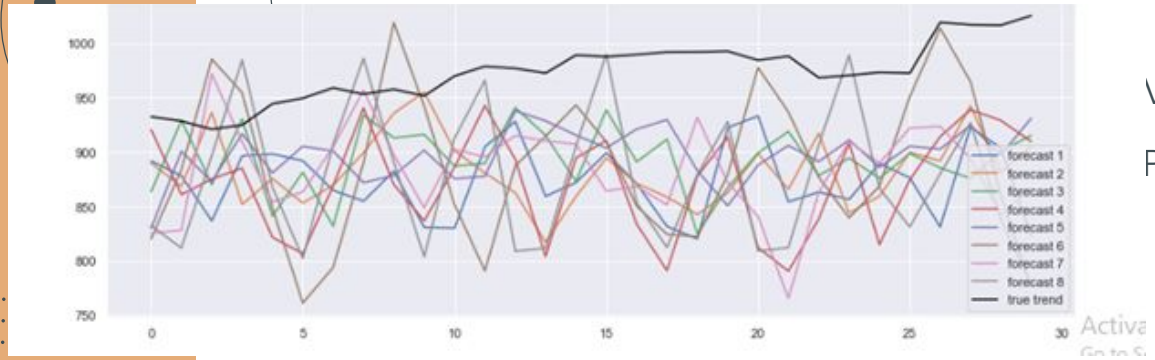
1. Historical Data Collection
2. Data Processing
    a. Extraction of Closing Stock Prices
    b. Normalisation
    c. Train Test Split(220+30)
3. Models Considered -
    a. RNN
    b. LSTM
    c. GRU
    d. CNN
4. Variations of these models -
    a. Vanilla
    b. Bidirectional
    c. 2-Path
    d. Sequence to Sequence
    e. Bidirectional Sequence to Sequence

Classical Approach Experimental Steps

1. Model Construction
    a. Hyperparameters: num_layers = 1,size_layer = 128, timestamp = 5, epoch = 10, dropout_rate = 0.8,future_day = test_size, learning_rate = 0.01, test_size=30
2. Training
    a. 220 Days
    b. Batch Size=5 Data Points
3. Testing
    a. Test Size=30 Days
4. Running Simulations
5. Calculating accuracy
6. Plotting Graphs

Predicted Results

RNN



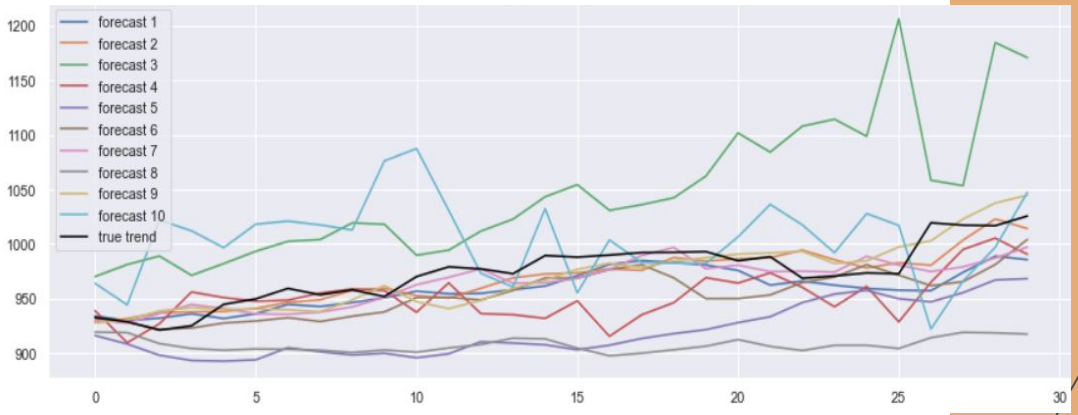
Volatility accuracy=53.3337%

Price accuracy = 91.4686%

Volatility accuracy=56.4%

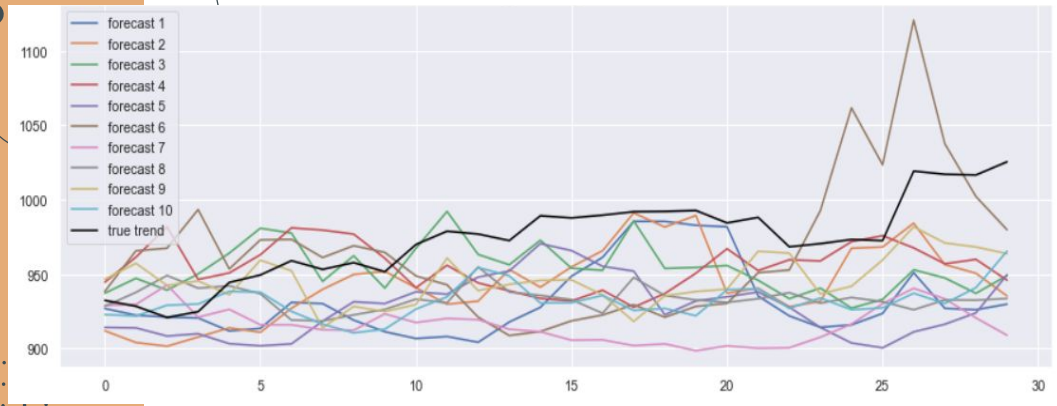
Price accuracy = 92.56%

GRU



Predicted Results

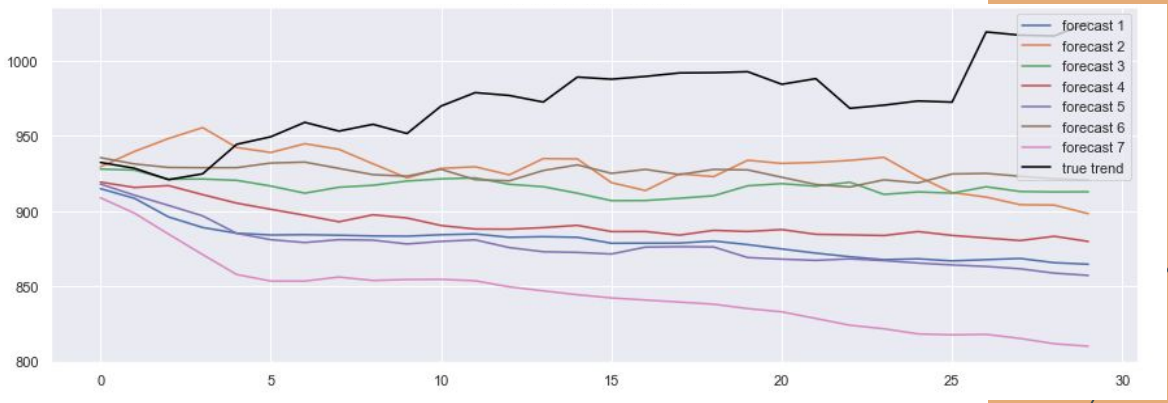
LSTM



Volatility accuracy=57.8%

Price accuracy = 94.98%

CNN



Volatility accuracy=55.2346%

Price accuracy = 90.7365%

Conclusion

Stock Price Accuracy is calculated using MSE(Mean Squared Error)
Volatility accuracy is the number of days we correctly predict direction of stock trend.

Model	Volatility accuracy	Price accuracy
RNN	53.34%	91.5%
GRU	56.4%	92.56%
LSTM	57.8%	**94.9%**
CNN	55.24%	90.74%

We observed that LSTM models performed slightly better on average than the others for predicting stock market price and volatility.

Stock volatility prediction using Sentiment Analysis

News Article Generation

Web Scraper made in Python is used to get the list of links of news articles. This uses the python module ‘scrapy’. We use the website ‘reuters.com’ for this purpose. For each day we have a corresponding page on the website which lists all the news articles of the given company on that day. We use scrapy to extract the links of the news articles from the html anchor tags on the page. Then we output it in a csv file with two columns, the date and url of the news articles.

	A	B
1	date	url
2	11112012	/article/apple-htc-settlement/update-1-apple-and-htc-settle-global-patent-battle-idUSL1E8MB05520121111
3	11112012	/article/apple-htc-settlement/apple-and-htc-settle-global-patent-war-idUSL1E8MB04M20121111
4	11122012	/article/us-htc-stx/htc-shares-jump-after-settles-patent-issues-with-apple-idUSBRE8AB01N20121112
5	11122012	/article/htc-stx/update-1-htc-shares-jump-after-settles-patent-issues-with-apple-idUSL3E8MC0AD20121112
6	11122012	/article/press-digest-financial-tiems-nov-12/press-digest-financial-times-nov-12-idUSL5E8MC00320121112
7	11122012	/article/apple-htc-shares/shares-of-htc-limit-up-after-settling-patent-issues-with-apple-idUST8E8L801J20121112
8	11012012	/article/apple-ipad-components/apples-ipad-mini-includes-lcd-display-driver-from-rival-samsung-idUSL1E8M13JJ20121101
9	11012012	/article/us-apple-execs/apples-cook-fields-his-a-team-before-a-wary-wall-street-idUSBRE8A000W20121101
10	11012012	/article/apple-execs/apples-cook-fields-his-a-team-before-a-wary-street-idUSL1E8LV9E920121101
11	11132012	/article/htc-verizon/verizon-to-sell-htcs-droid-dna-smartphone-as-holiday-flagship-idUSL1E8MD5C020121113
12	11132012	/article/us-apple-italy-antitrust/apple-stops-selling-customer-protection-plan-at-italian-shops-idUSBRE8AC0VX20121113
13	11132012	/article/apple-italy-antitrust/apple-stops-selling-customer-protection-plan-at-italian-shops-idUSL5E8MDB8620121113
14	11132012	/article/htc-shares/update-1-taiwan-bourse-says-looking-into-unusual-htc-share-move-idUSL3E8MD0LW20121113
15	11132012	/article/htc-shares/taiwan-bourse-says-looking-into-unusual-htc-share-move-idUST8E8L801O20121113
16	11162012	/article/apple-patents/update-1-samsung-goes-after-htc-deal-to-undercut-apple-idUSL1E8MGDHW20121116
17	11162012	/article/us-usa-apple-stock/high-flying-apple-falls-to-earth-as-investors-fret-over-taxes-idUSBRE8AF18S20121116
18	11162012	/article/apple-patents/samsung-goes-after-htc-deal-to-undercut-apple-filing-idUSL1E8ME0QJ20121116
19	11162012	/article/usa-apple-stock/high-flying-apple-falls-to-earth-as-investors-fret-over-taxes-idUSL1E8MG9IY20121116
20	11162012	/article/us-apple-samsung/apple-samsung-allowed-to-add-products-in-u-s-patent-lawsuit-idUSBRE8AF05A20121116

Datasets Used

- News articles from Reuters.com on particular stock ranging from 2012 - 2018
- Historical Stock Prices from Yahoofinance.com

Sentiment Score Generation

The output csv file from Step A is used as an input in this step. Here we use the 'Rosette' API for text analysis for the contents for each article outputting a confidence label corresponding to each article.

The confidence column presents the probability/confidence with which the label is decided for that specific date Similarly confidence is the probability for the label while entity_confidence is probability for entity_labeld its share. The column confidence can range from -1: very negative outlook to 1: very positive outlook.

	date	label	confidence	entity-label	entity-confidence
2	1052012	pos	0.54460747	pos	0.57113739
3	1062012	neu	0.99008939		
4	1062012	neu	0.98103904	pos	0.49121883
5	1052012	neg	0.62086973	pos	0.57113739
6	1112012	neg	0.57316798		
7	1132012	neg	0.49429478	neu	0.40339307
8	1172012	neu	0.81733411	neu	0.97721458

Calculating Bullishness and combining with prices

Using the sentiment score calculated for each article in the previous step, We calculate the bullishness score of the selected stock on a per day basis.

$$Bullishness\ score = \frac{sum\ of\ confidence\ scores}{Number\ of\ articles}$$

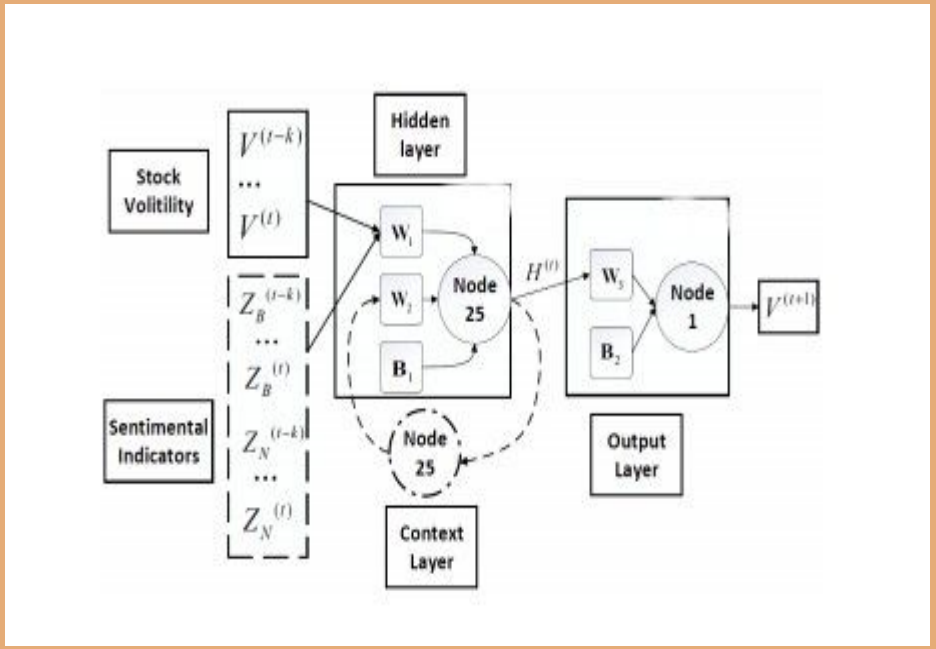
Bullishness score represents the average perspective of the particular stock on the particular date. Afterwards, we combine it with the stock price and sort it in chronological order.

In case, the sentiment score of a day is not available, mark the sentiment score as the score of the latest previous day available.

	A	B	C
1	date	opinion	price
2	20121112	-0.264832311428571	226.470001
3	20121113	-0.264832311428571	226.600006
4	20121114	-0.1825263075	222.949997
5	20121115	-0.1825263075	220.600006
6	20121116	-0.1825263075	225.229996
7	20121119	-0.1825263075	229.710007
8	20121120	-0.1825263075	233.779999
9	20121121	-0.1825263075	238.029999
10	20121123	-0.1825263075	239.880005
11	20121126	0	243.619995
12	20121127	-0.121427378	243.399994
13	20121128	-0.121427378	247.110001

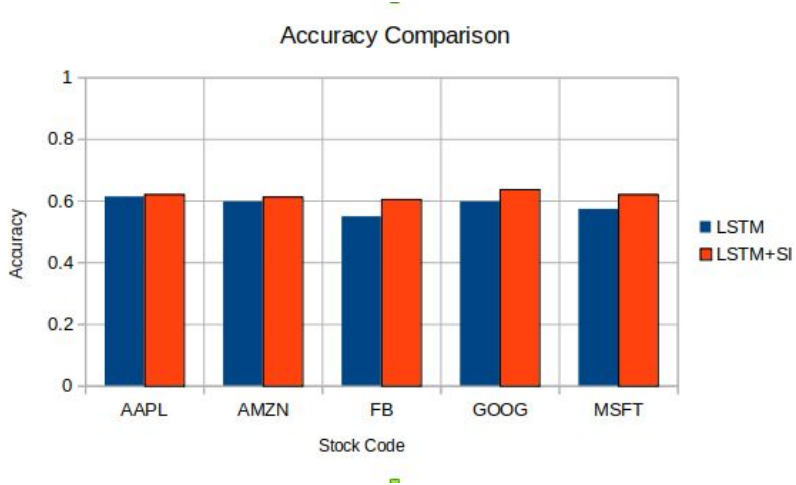
LSTM Model

For training our model, the sentiment score goes z score normalization followed by min-max normalization. Then, we construct a 2 layer LSTM followed by 2 Dense layers. Input of which will be previous k days stock price and it's corresponding sentiment score where k is the window size. The output is the prediction of stock price. Hyperparameters of this model are as follows batch_size=128, epochs=200, validation_split=0, verbose=0.



Result and Conclusion

Accuracy and the best k for RNN+EMM and RNN					
	AAPL	AMZN	FB	GOOG	MSFT
LSTM	0.6129	0.5968	0.5484	0.5968	0.5726
LSTM+SI	0.621	0.6129	0.6048	0.6371	0.621
k(LSTM)	4	7	6	7	6
k(LSTM+SI)	4	9	7	7	5



We used RMSProp optimizer to minimize mean square error. Our objective is to maximise stock volatility prediction accuracy.
After completing the prediction, we can clearly observe that the results using sentiment analysis offer marginally better results as compared to predicting without them. Hence, we conclude that analyzing sentiments in order to predict the volatility in the stock market is not worth the effort.

Challenges and Learnings

Challenges

- Historical news article collection was a major bottleneck.
- Selection of sentiment score generation method from news articles.
- Training and optimizing large number of models.

Learnings

- Familiarizing ourselves with stock market and it's terminology.
- Learning and implementing different variations of the neural networks.
- Gaining first hand knowledge of using various tools like Scrappy, Rosette etc

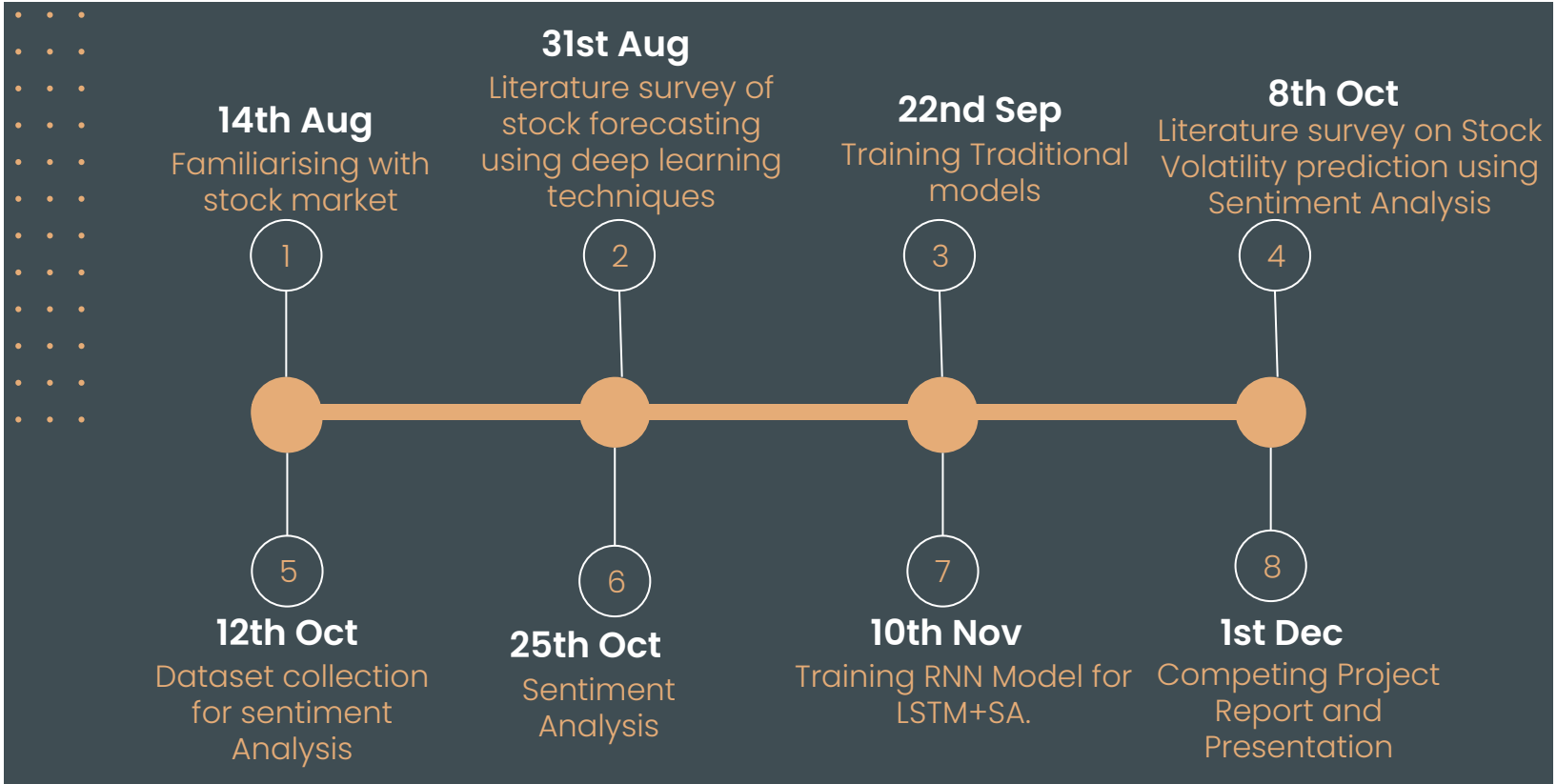
Future Scope

- Building the sentiment score module from scratch instead of outsourcing it to Rosette.
- Merging multiple social media sources (like Twitter, reddit) and unconventional sources (like Youtube) for sentiment analysis.
- Creating a mechanism that only recommend when there is a noticeable sentiment undercurrent in the market for stronger results.

List of Papers Referred

S. No	Research Paper	Author(s)
1.	Stock Prices Prediction using Deep Learning Models (2019)	*Jialin Liu, Fei Chao, Yu-Chen Lin and Chih-Min Lin*
2.	Neural networks for stock price prediction (2018)	*Ren-Jie Han, Yu-Long Zhou and Yue-Gang Song*
3.	Stock Trend Prediction using News Sentiment Analysis	*Kalyani Joshi , Prof. Bharathi H. N. and Prof. Jyothi Rao*
4.	Sentiment Analysis of Twitter Data for Predicting Stock Market Movements (2016)	*Venkata Sasank Pagolu*
5.	Stock Volatility Prediction Using Recurrent Neural Networks with Sentiment Analysis	*Yifan Liu, Zengchang Qin, Pengyu Li and Tao Wan*
6.	Market Trend Prediction using Sentiment Analysis: Lessons Learned and Paths Forward	*Andrius Mudinas and Dell Zhang*
7.	Stock Forecasting using M-Band Wavelet-Based SVR and RNN-LSTMs Models	*Xiaodi Wang and Abdul Hasib Rahimyar*
8.	A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing	*Sidra Mehtab and Jaydip Sen*

Timeline


14th Aug
Familiarising with stock market
1
31st Aug
Literature survey of stock forecasting using deep learning techniques
2
22nd Sep
Training Traditional models
3
8th Oct
Literature survey on Stock Volatility prediction using Sentiment Analysis
4
5
12th Oct
Dataset collection for sentiment Analysis
6
25th Oct
Sentiment Analysis
7
10th Nov
Training RNN Model for LSTM+SA.
8
1st Dec
Competing Project Report and Presentation

PLAGIARISM REPORT

Turnitin Originality Report

Stock Price Forecasting by Harshit Mahajan

From btp_grp (Research)

Processed on 30-Dec-2020 18:51 IST
ID: 1482078877
Word Count: 8050

Similarity Index	Similarity by Source	
7%	Internet Sources:	2%
	Publications:	1%
	Student Papers:	4%

sources:

1. 2% match (student papers from 28-Dec-2020)
   Submitted to Netaji Subhas Institute of Technology on 2020-12-28

2. 2% match (student papers from 27-May-2019)
   Submitted to Netaji Subhas Institute of Technology on 2019-05-27

3. 1% match (Internet from 31-Oct-2019)
   http://export.arxiv.org/pdf/1909.12227

4. 1% match (Internet from 15-Jan-2019)
   https://arxiv.org/pdf/1805.11317.pdf

5. 1% match (publications)
   Yi Luo, Zhuo Chen, Takuya Yoshioka, "Dual-Path RNN: Efficient Long Sequence Modeling for Time-Domain Single-Channel Speech Separation", ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020

6. 1% match (Internet from 12-Jun-2020)
   https://capital.com/stock-market-prediction-definition

Thanks!

We'll be happy to answer any questions!!!