Contents

[1 Introduction 2](#_Toc532206776)

[Objective 2](#_Toc532206777)

[Tools 3](#_Toc532206778)

[Structure of the paper 3](#_Toc532206779)

[2 Background 3](#_Toc532206780)

[Sentiment Analysis (Twitter) 3](#_Toc532206781)

[Technical Analysis 3](#_Toc532206782)

[Fundamental Analysis 3](#_Toc532206783)

[LSTM best model for stock prediction 3](#_Toc532206784)

[Previous Work / Literature Review 3](#_Toc532206785)

[3 Data Collection 4](#_Toc532206786)

[Time Series Data Collection (real-time, offline) 4](#_Toc532206787)

[Time Series Data Preprocessing 4](#_Toc532206788)

[4 System Design 4](#_Toc532206789)

[Object Oriented Programming 4](#_Toc532206790)

[LSTM Model 4](#_Toc532206791)

[Univariate one-step forecasting LSTM 4](#_Toc532206792)

[Univariate multi-step forecasting LSTM 4](#_Toc532206793)

[Multivariate one-step forecasting LSTM 4](#_Toc532206794)

[Multivariate multi-step forecasting LSTM 4](#_Toc532206795)

[Realtime 4](#_Toc532206796)

[5 Stock price prediction Implementation 4](#_Toc532206797)

[Parameter optimisation 4](#_Toc532206798)

[Training 5](#_Toc532206799)

[Data Preprocessing for Supervised Learning in LSTM 5](#_Toc532206800)

[Correlation with technical indicators 5](#_Toc532206801)

[Correlation with fundamental indicators 5](#_Toc532206802)

[Prediction Visualisation 5](#_Toc532206803)

[One step Forecasting 5](#_Toc532206804)

[Univariate Time Series 5](#_Toc532206805)

[Multivariate Time Series 5](#_Toc532206806)

[Multi-step result 5](#_Toc532206807)

[Univariate Time Series 5](#_Toc532206808)

[Multivariate Time Series 5](#_Toc532206809)

[6 Evaluation 5](#_Toc532206810)

[Metrics 5](#_Toc532206811)

[APE (absolute percentage error) 5](#_Toc532206812)

[AAE (average percentage error) 5](#_Toc532206813)

[ARPE (average relative percentage error) 5](#_Toc532206814)

[RMSE (root mean squared error) 5](#_Toc532206815)

[MCC (Matthews correlation coefficient) 5](#_Toc532206816)

[One step Forecasting 5](#_Toc532206817)

[Univariate Time Series 5](#_Toc532206818)

[Multivariate Time Series 5](#_Toc532206819)

[Multi-step result 5](#_Toc532206820)

[Univariate Time Series 5](#_Toc532206821)

[Multivariate Time Series 5](#_Toc532206822)

[Comparing baseline result using various metric 5](#_Toc532206823)

[7 Conclusions and Future Work 5](#_Toc532206824)

[Fundamental analysis taken into account 5](#_Toc532206825)

[Online self-learning model 5](#_Toc532206826)

[Tutorials 6](#_Toc532206827)

Title and Abstract

# 1 Introduction

## Objective

1. Predict quarterly result from social media, technical and fundamental analysis, industry macroeconomic data, and news on the same day
2. Allows user to “watch” specific stock’s general sentiments and alerts of important news which may affect stock price
3. Self-learning model for each “Watch” stock after each quarter
4. Tracking earning release?
5. Analyst consensus price target performance, sentiment on stocks buy neutral and sell. Train a ml model by checking if the price target meet expectation and use it to predict the chances of the stock meeting each experts. Revenue estimates, EPS, etc… and build a profile for each analyst
6. Chance events/news to affect stock index and individual stocks overall.
7. Candlestick Patterns
8. If certain (undervalued) companies outperformed analyst expectation in the past
9. Attractive industry to invest in?
10. Industry-wide performance, industry outperformers and underperformers.
11. Industry growth in market cap?
12. Simulated Trading return?

## Tools

## Structure of the paper

# 2 Background

## Sentiment Analysis (Twitter)

## Technical Analysis

## Fundamental Analysis

## LSTM best model for stock prediction

## Previous Work / Literature Review

LSTM vs Feedforward network:

Feedforward network’s outputs are independent of each other thus cannot capture the dependencies

They might have optimised parameters

* <http://www.cs.cmu.edu/~bdhingra/papers/stock_hmm.pdf>
  + HMM
  + ARIMA (AutoRegressive Integrated Moving Average)
  + ANN (artificial neural network)
  + IBM, APPLE, TATA STEELS, DELL
* Deep Learning
* Convolutional NN
* Unsupervised
* Supervised
* Logistic regression
* Linear Regression 2) Stochastic Gradient Descent (SGD) 3) Support Vector Regression (SVR)L
  + S&P500 - <https://pdfs.semanticscholar.org/0096/f7f6b0724c5839163b0e851b12b32f8ec908.pdf>
* Radial basis function SVC:
* Polynomial SVC
* LSTM
  + <https://arxiv.org/pdf/1603.07893.pdf> Google
* Vector Auto Regression

# 3 Data Collection

## Time Series Data Collection (real-time, offline)

Freemium

* Alpha Vantage
  + [https://www.alphavantage.co/query?function=RSI&symbol=MSFT&interval=daily&time\_period=10&series\_type=open&apikey=**3OMS720IM6CRC3SV**](https://www.alphavantage.co/query?function=RSI&symbol=MSFT&interval=daily&time_period=10&series_type=open&apikey=3OMS720IM6CRC3SV)
  + **Github wrapper**
* IEX Trading
* SimFin

Premium

* Zacks (1000+$)
* Intrinio (200+$)
* XIgnite (1000+$)

## Time Series Data Preprocessing

Dates

No outliers due to reliability of the api

# 4 System Design

## Object Oriented Programming

## One-step forecasting LSTM Model

### Baseline one-step forecasting model

### Univariate one-step forecasting LSTM

### Multivariate one-step forecasting LSTM

## Multi-step forecasting LSTM Model

### Baseline multi-step forecasting model

### Univariate multi-step forecasting LSTM

### Multivariate multi-step forecasting LSTM

## Realtime

# 5 Stock price prediction Implementation

## Parameter optimisation

Hypothesis, more data = better, test it

## Training

## Data Preprocessing for Supervised Learning in LSTM

## Correlation with technical indicators

## Correlation with fundamental indicators

## Prediction Visualisation

## Robust result

Train 10 models and get the average of the result

## One step Forecasting

### Univariate Time Series

### Multivariate Time Series

## Multi-step result

### Univariate Time Series

### Multivariate Time Series

# 6 Evaluation

## Metrics

### APE (absolute percentage error)

### AAE (average percentage error)

### ARPE (average relative percentage error)

### RMSE (root mean squared error)

### MCC (Matthews correlation coefficient)

Training and Testing set

## One step Forecasting

### Univariate Time Series

### Multivariate Time Series

## Multi-step result

### Univariate Time Series

### Multivariate Time Series

## Comparing baseline result using various metric

* Univariate Persistence Model Forecast
* Multivariate Persistance Model Forecast
* Comparison with existing results from other papers (HMM, ARIMA, ANN)

# 7 Conclusions and Future Work

## Fundamental analysis taken into account

## Online self-learning model

[Alpha Vantage](https://www.alphavantage.co/documentation/)

[IEX](https://iextrading.com/developer/docs)

# Attributions

Holidays <https://pypi.org/project/holidays/> (we’re using American stocks)

Adding business days: <https://stackoverflow.com/questions/12691551/add-n-business-days-to-a-given-date-ignoring-holidays-and-weekends-in-python/23352801>

Alpha Vantage Wrapper: <https://github.com/RomelTorres/alpha_vantage>

Degiro: <https://github.com/pladaria/degiro>

Backbone for one-step forecasting LSTM <https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/>

Multi-step forecasting LSTM: <https://machinelearningmastery.com/multi-step-time-series-forecasting-long-short-term-memory-networks-python/>

# Tutorials

<https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>

<https://www.quora.com/Which-regression-model-is-best-for-predicting-forecasting-stock-prices>