Contents

[1 Introduction 2](#_Toc531883333)

[Objective 2](#_Toc531883334)

[Tools 2](#_Toc531883335)

[Structure of the paper 2](#_Toc531883336)

[2 Background 2](#_Toc531883337)

[Sentiment Analysis (Twitter) 2](#_Toc531883338)

[Technical Analysis 2](#_Toc531883339)

[Fundamental Analysis 2](#_Toc531883340)

[LSTM best model for stock prediction 2](#_Toc531883341)

[Previous Work / Literature Review 2](#_Toc531883342)

[3 Data Collection 3](#_Toc531883343)

[Time Data Collection (real-time, offline) 3](#_Toc531883344)

[Time Data Preprocessing 3](#_Toc531883345)

[Object Oriented Programming 3](#_Toc531883346)

[LSTM vs Logistic Regression 3](#_Toc531883347)

[4 Stock price prediction Implementation 3](#_Toc531883348)

[LSTM Model 3](#_Toc531883349)

[Realtime 4](#_Toc531883350)

[Parameter optimisation 4](#_Toc531883351)

[Training 4](#_Toc531883352)

[Data Preprocessing for Supervised Learning in LSTM 4](#_Toc531883353)

[Correlation with technical indicators 4](#_Toc531883354)

[Correlation with fundamental indicators 4](#_Toc531883355)

[Prediction Visualisation 4](#_Toc531883356)

[6 Evaluation 4](#_Toc531883357)

[7 Conclusions and Future Work 4](#_Toc531883358)

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)
* 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 Data Collection (real-time, offline)

Freemium

* Alpha Vantage
* IEX Trading
* SimFin

Premium

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

## Time Data Preprocessing

Dates

## Object Oriented Programming

## LSTM vs Logistic Regression

# 4 Stock price prediction Implementation

## LSTM Model

Univariate one-step forecasting LSTM

Univariate multi-step forecasting LSTM

Multivariate one-step forecasting LSTM

Multivariate multi-step forecasting LSTM

## Realtime

## Parameter optimisation

## Training

## Data Preprocessing for Supervised Learning in LSTM

## Correlation with technical indicators

## Correlation with fundamental indicators

## Prediction Visualisation

# 6 Evaluation

Metrics

* APE
* AAE
* ARPE
* RMSE

5.0 Training and Testing set

5.1 Baseline result

* Univariate Persistence Model Forecast

5.1.1 Univariate Time Series

5.1.2 Multivariate Time Series

5.2 One step Forecasting

5.2.1 Univariate Time Series

5.2.2 Multivariate Time Series

5.3 Multi-step result

5.3.1 Univariate Time Series

5.3.2 Multivariate Time Series

5.4 Comparison with existing results (HMM, ARIMA, ANN)

# 7 Conclusions and Future Work

6.1 Fundamental analysis taken into account

6.2 Online self-learning model

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

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

SimFin