**Machine Learning for Long Term Stock Price Prediction**

“submitted towards partial fulfilment of the criteria for award of PGPBA by GLIM”

Submitted by: Sahil Makkar, Utkarsh Kulshrestha,

Ashwini Murthy, Deepti Ramani,

Pavan Kumar Thatha, Showbhik

Group No: 4

Batch: PGPBA-BI 2017, Section A

Project Mentor: Neelesh Singh



GREAT LAKES INSITUTE OF MANAGEMENT, BANGALORE

# **Acknowledgements**

We wish to place on record our deep appreciation for the guidance and help provided to us by our Mentor Mr. Neelesh Singh from Great Lakes Institute of Management for guiding us in completing this project on time.

We would also like to place on record our appreciation for the guidance provided by Dr. P K Vishwanathan for giving us valuable feedback and being a source of inspiration in helping us to work on this project.

Last but not the least we wish to thank Project Office, Bangalore for their support.

We certify that the work done by us for conceptualizing and completing this project is original and authentic.

Date: 25 November 2017

Place: Bangalore

Sahil Makkar

UtkarshKulshrestha

Ashwini Murthy

Deepthi Ramani

Pavan Thatha

Showbhik

**GREAT LAKES INSTITUTE OF MANAGEMENT**

Postgraduate Program Business Analytics and Business Intelligence



***Certificate***

This is to certify that the PGPBA Capstone Project titled “***Machine Learning for Long Term Stock Price Prediction***” is a bonafide record of the Project work carried out by *Sahil Makkar, Utkarsh Kulshrestha, Ashwini Murthy, Deepthi Ramani, Pavan Thatha and Showbhik* in fulfilment of requirements for the award of PGPBA-BI in Great Lakes Institute of Management.

November 2017

**Neelesh Singh** **Dr. P K Vishwanathan**

Project Mentor Program Director

Date: November 2017

Place: Bangalore

# **Abstract**

Details must include:

* Techniques: (from predictive modelling and data mining)
* Tools: (R, SAS, Hadoop etc)
* Domain: Domain Analytics name

# **Table ofContents**

[**Acknowledgements** 2](#_Toc499588559)

[**Abstract** 4](#_Toc499588560)

[**Table ofContents** 5](#_Toc499588561)

[**List of Tables** 6](#_Toc499588562)

[**List of Figures** 7](#_Toc499588563)

[**Chapter 1** 8](#_Toc499588564)

[**Project Objective and Business Problem** 8](#_Toc499588565)

[**Data Sources** 8](#_Toc499588566)

[**Statistical Tools and Techniques Used** 9](#_Toc499588567)

[**Chapter 2** 10](#_Toc499588568)

[**Literature Overview** 10](#_Toc499588569)

[**Importance of Stock Price Prediction** 11](#_Toc499588570)

[**Important Terminologies in Stock Market** 11](#_Toc499588571)

[**Financial Ratios** 12](#_Toc499588572)

[**Chapter 3** 13](#_Toc499588573)

[**Data under Study** 13](#_Toc499588574)

[**Analytical Approach** 13](#_Toc499588575)

[**Exploratory Data Analysis** 13](#_Toc499588576)

[**Chapter 4** 14](#_Toc499588577)

[**Sentiment Analysis** 15](#_Toc499588578)

[**Predictive Analytics** 15](#_Toc499588579)

[**Chapter 5** 19](#_Toc499588580)

[**Recommendations** 19](#_Toc499588581)

[**Conclusions** 20](#_Toc499588582)

[**Bibliography** 21](#_Toc499588583)

[**Annexure** 22](#_Toc499588584)

# **List of Tables**

# **List of Figures**

# **Chapter 1**

## **Introduction**

The stock market is essentially a non-linear, nonparametric system that is extremely hard to model with any reasonable accuracy. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. To achieve those objectives, and according to some research used the techniques of fundamental analysis, where trading rules are developed based on the information associated with macroeconomics, industry, and company. The authors of some of the research said that fundamental analysis assumes that the price of a stock depends on its intrinsic value and expected return on investment. Analysing the company’s operations and the market in which the company is operating can do this. Consequently, the stock price can be predicted reasonably well. Most people believe that fundamental analysis is a good method only on a long-term basis. However, for short- and medium-term speculations, fundamental analysis is generally not suitable. Some other research used the techniques of technical analysis, in which trading rules were developed based on the historical data of stock trading price and volume. Technical analysis as illustrated in and refers to the various methods that aim to predict future price movements using past stock prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data. Thus, it is assumed that price trends and patterns exist that can be identified and utilized for profit. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid.

However, now data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area. Artificial neural networks can be trained by using an appropriate learning algorithm. Following the assumption of technical analysis that patterns exist in price data, it is possible in principle to use data mining techniques to discover these patterns in an automated manner. Once these patterns have been discovered, future prices can be predicted.

In this project we will Introduce a framework in which we integrate user predictions into the current machine learning algorithm using public historical data & text mining to improve our results. Following the assumption of technical analysis that patterns exist in price data, it is possible in principle to use data mining techniques to discover these patterns in an automated manner. Once these patterns have been discovered, future prices can be predicted.

Fundamental Analysis:

Fundamental Analysis is a method of evaluating a security in an attempt to measure its intrinsic value, by examining related economic, financial and other qualitative and quantitative factors. Fundamental analysts study anything that can affect the security’s value, including macroeconomic factors such as the overall economy and industry conditions and microeconomic factors such as financial conditions and company management. The end goal of fundamental analysis is to produce a quantitative value that an investor can compare with a security’s current price, thus indicating whether security is undervalued or overvalued.

Fundamental analysis involves the process of:

* Looking at the basics or fundamental quantitative and financial environment of a business.
* Examining key ratios to determine financial health
* Estimating the value of the stock

For stocks and equity instruments, this method uses revenues, earnings, future growth, ROE, profit margins and other data to determine a company's underlying value and potential for future growth. In terms of stocks, fundamental analysis focuses on the financial statements of the company being evaluated.

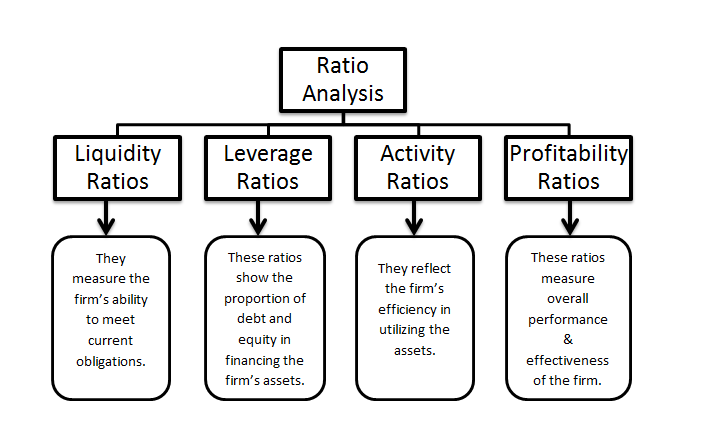


The goal of analyzing a company’s fundamentals is to find a stock’s intrinsic value, a term which gives what a stock in really worth – as opposed to a value at which it is being traded. If the intrinsic value is more than the current share price, our analysis is showing that the stock is worth more than its price and that it makes sense to buy this stock. In terms of stocks, fundamental analysis focuses on the financial statements of the company being evaluated.

Financial Ratios:

The main objective of a financial a report is to provide information on a company’s performance to the internal and external users to take decisions. From the perspective of information economics, accounting and financial reporting play a vital role in an efficient capital market. One of the most common ways of assessing the relative values of stocks among practitioners is to compare the numbers listed in financial statements by using financial ratios. The main advantage of using financial ratios instead of amounts from the income statement is that they are independent of the size of the company. The comparison of financial ratios is used to assess companies’ financial condition, operations and attractiveness as an investment.

Financial ratios defined as the numerical value created from two or more values taken from a company’s financial statements i.e. its balance sheet, income statement or statement of cash flow. Typically, financial ratios are presented as a quantified metric in the form of a percentage, multiple or a ratio which aims to evaluate the financial, operational performance and competitiveness of a company.



**Liquidity Ratios**: The liquidity focuses on a firm’s ability (current assets and current liabilities) to meet its short-term debt obligations. In other words, it lets you know the resources available for a firm to use in order to pay its current obligation and expenses. If a company cannot maintain a short-term debt-paying ability, it will not be able to maintain a long-term debt-paying ability, nor will it be able to satisfy its stockholders. The main ratio in this group is the current ratio and acid test ratio.

Current Ratio is the mort widely used of all analytical devices based on the balance sheet. It establishes the relationship between Current assets & Current liabilities.

Current Ratio = Current Assets / Current Liabilities

Ideal Ratio: 2:1

High Ratio indicates under trading and over capitalization.

Low Ratio indicates over trading and under capitalization.

**Leverage Ratio**: A leverage ratio is any one of several financial measurements that look at how much capital comes in the form of debt (loans), or assesses the ability of a company to meet its financial obligations. The leverage ratio is important given that companies rely on a mixture of equity and debt to finance their operations, and knowing the amount of debt held by a company is useful in evaluating whether it can pay its debts off as they come due.

Too much debt can be dangerous for a company and its investors. However, if a company's operations can generate a higher rate of return than the interest rate on its loans, then the debt is helping to fuel growth in profits. Nonetheless, uncontrolled debt levels can lead to credit downgrades or worse. On the other hand, too few debts can also raise questions. A reluctance or inability to borrow may be a sign that operating margins are simply too tight.

The most well-known financial leverage ratio is the debt-to-equity ratio.

D/E Ratio = Total Debt / Total Equity

A high debt/equity ratio generally indicates that a company has been aggressive in financing its growth with debt. This can result in volatile earnings as a result of the additional interest expense. If the company's interest expense grows too high, it may increase the company's chances of a default or bankruptcy. Typically, a D/E ratio greater than 2.0 indicates a risky scenario for an investorhowever this yardstick can vary by industry.

**Profitability Ratio:**

The profitability ratios group, also known as performance ratios, assesses the company ability to earn profits on sales, assets and equity, it measures the return earned on a company’s capital and the financial cushion relative to each dollar of sales, These are critical to determining the attractiveness of investing in company shares, and investors in using these ratios widely, much like the operational performance ratios, these ratios give users a good understanding of how well the company utilized its resources(assets) in generating profit and shareholder value. The long-term profitability of a company is vital for both the survivability of the company as well as the benefit received by shareholders.

# **Chapter 2**

## **Literature Overview**

## **Importance of Stock Price Prediction**

## **Important Terminologies in Stock Market**

## **Financial Ratios**

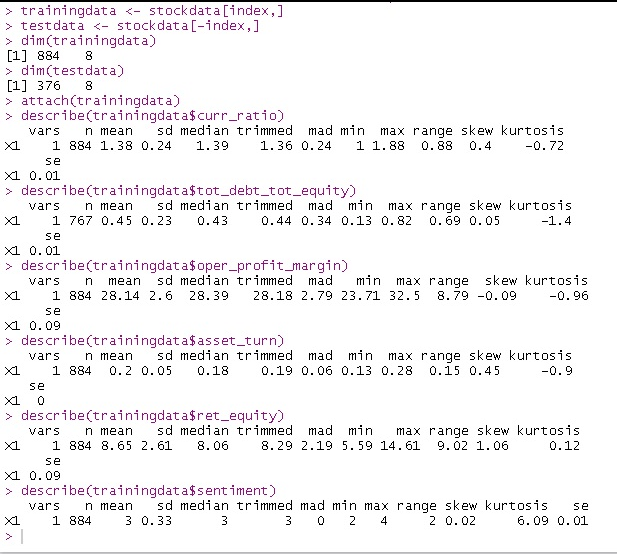
# **Chapter 3**

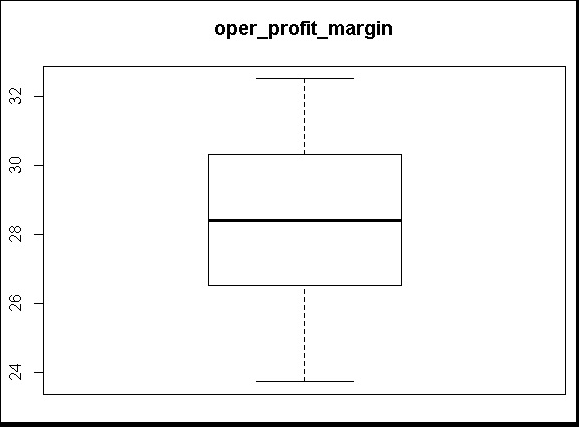
## **Data under Study**

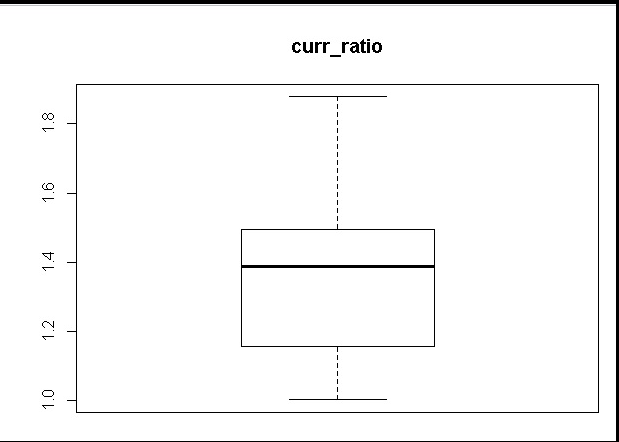
## **Analytical Approach**

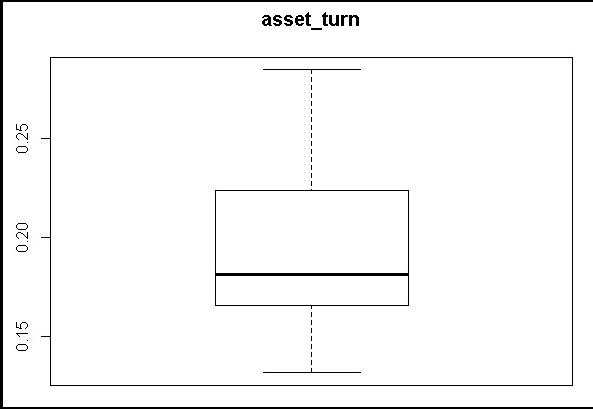
## **Exploratory Data Analysis**

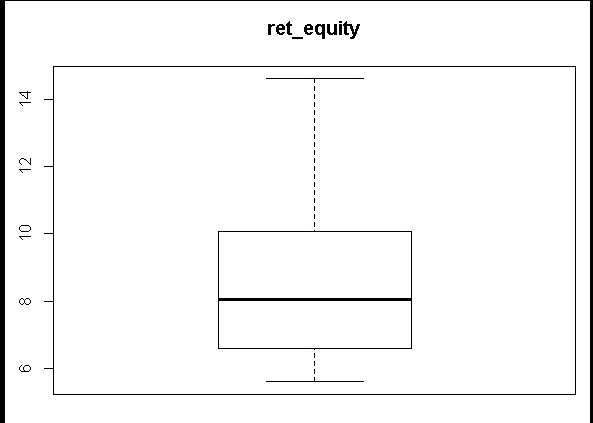
We have picked few sample stocks for EDA and here are the findings.

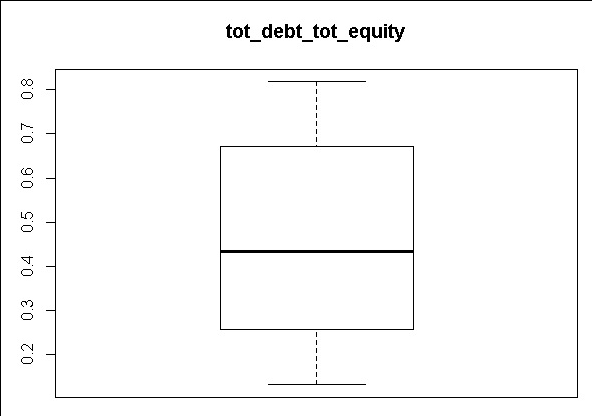


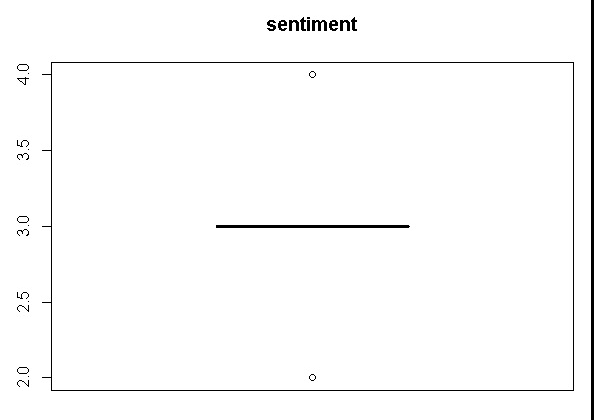


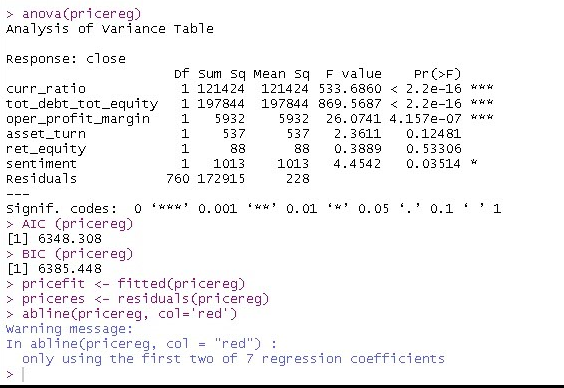












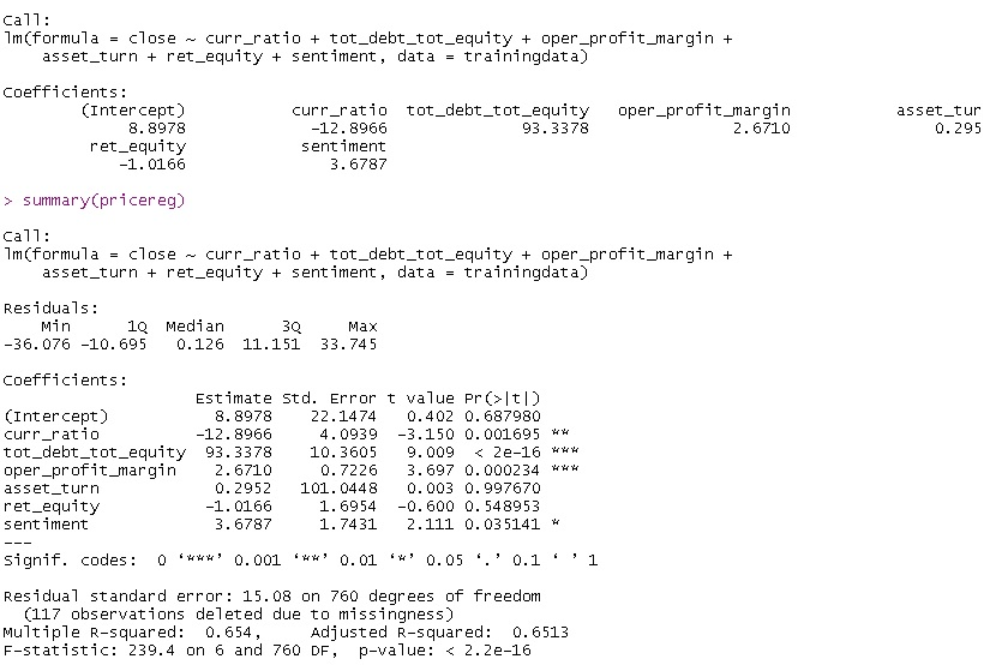
# **Chapter 4**

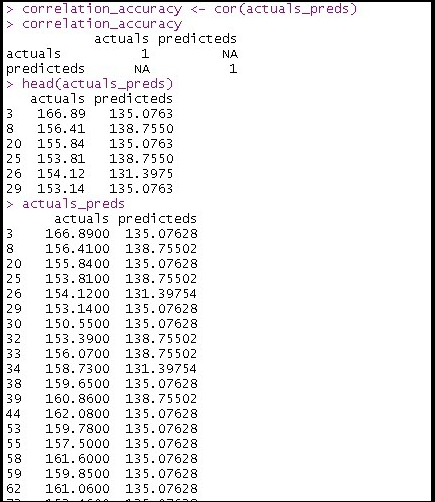
## **Sentiment Analysis**

## **Predictive Analytics**

In order to find a suitable technique for predicting the stock price accurately, we tried a pool of 14 prediction models (7 with regression and other 7 for Neural Networks)

Regression with Sentiment scores could n’t result a reasonable prediction model, here are the few sample screenshots





Neural Network model yielded a much better prediction model with the below sample outputs.

**Neural Model:**

We have multi layer model having two LSTM layers, followed by one feed forward layer and one softmax based output layer.

Layer (type) Output Shape Param #

=================================================================

lstm\_3 (LSTM) (None, 5, 128) 69120

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

lstm\_4 (LSTM) (None, 128) 131584

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dropout\_2 (Dropout) (None, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_3 (Dense) (None, 64) 8256

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_4 (Dense) (None, 3) 195

=================================================================

Total params: 209,155

Trainable params: 209,155

Non-trainable params: 0

Neural model classifies Stocks based on 5 days of sequence and predicting 5th day. It's multi feature sequence model.

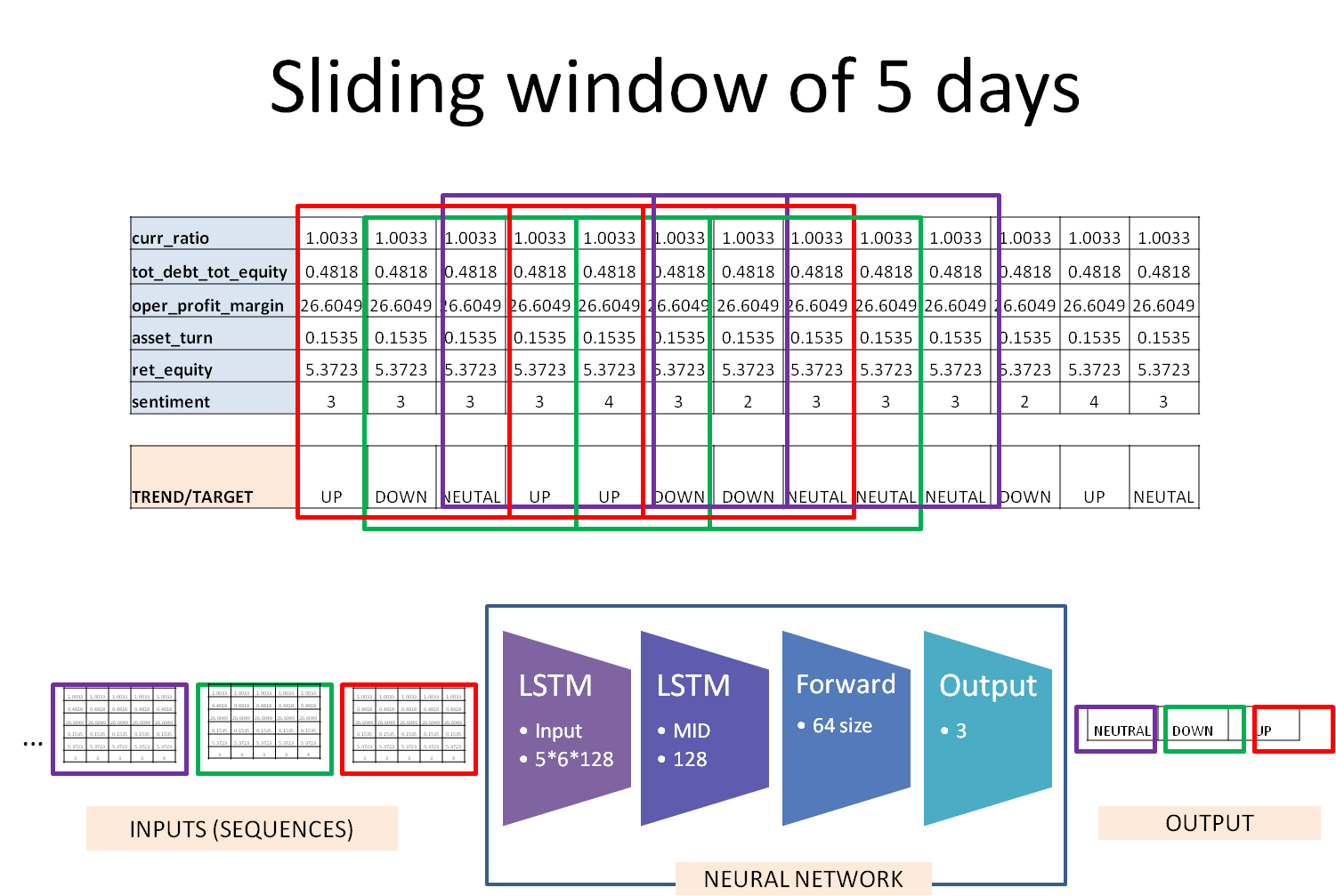
We have 3 classes :

1. Positive recommendation (where a day gain is more than 1%),
2. Negative recommendation (where a day loss is greater than 1%) and
3. Neutral recommendation for all other cases

**Approach:**

**Key features are:**

1. Inputs to ML algorithm (neural network) are formed with 5 days window to understand weekly impact.
2. 3 class neural network to predict : SELL/BUY/HOLD is done
3. Neural network has 3 layers, 2 LSTM to capture the sequence impact, one feed forward and then finally 3 class output classifier.
4. Regression based model because of limited data is not considered, but same model can be converted to regression to predict prices itself.



Given model gives 96% of accuracy in predictions. Out-sampling cases need to be handled better. Lag of data can help in better accuracy across classes.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# **Chapter 5**

## **Recommendations**

## **Conclusions**

# **Bibliography**

# **Annexure**