**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

Utkarsh Kulshrestha

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 of** **Contents**

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

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

[**Table of** **Contents** 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**

## **Project Objective and Business Problem**

The study will be conducted on multi-industry stocks. Approx. 10 stocks are selected from different Industries.

**Need of Study** - Stock market has become one of the major components of economy not only in the developed countries but also in the developing countries. Making decision in stock market is not really easy because a lot of factors are involved with every choice we make. Therefore, a lot of analysis is required to make an optimal move on stock market which may involve price trend, market’s nature, company’s stability, economy, different news and rumors about stocks etc. The objective of the study is to extract fundamental information, technical factors (Inflation, Economy, Market strength, Demographics), Market sentiments, news and use them to analyze and forecast the stock market from the company’s investor viewpoint.

**Objectives:**

* To study the fundamental, technical & sentiment analysis of a universe of few US companies.
* To build a Deep/Machine Learning model for stock prices prediction.
* Text mining to be done on Twitter, business news website scrapping and a few other sources.
* Fundamental data extraction.
* Automation of fundamental extraction.
* Applying fundaments on historical data, sentiment analysis and building model to forecast the stock prices.

## **Data Sources**

Below are few data sources that have been used.

* ***Stock Data***: NYSE and NASDAQ, Quandl API
* ***News Data***: Money control, News websites.
* ***Social Media Data***: Twitter and few others.

## **Statistical Tools and Techniques Used**

**Algorithm:**

Ratio chosen and sentiments data would be added in sequence form to neural network to predict price for next even. Next event granularity duration can be from a week to quarter to year as well.

**Techniques:**

* Twitter stream listening
* Ratio analysis and calculation from historical data.
* Labeling of financial tweets.
* Sentiment analysis of tweets.
* Deep/Machine learning to predict stock prices.

**Tools:**

* Python
* TensorFlow
* Keras
* R Studio
* Tableau

# **Chapter 2**

## **Literature Overview**

Stock market is very versatile and fluctuates with time. Technically, this fluctuation can be explained by fundamentals of organization, industry and economy. Long term stock price is determined by various financial analyst using amalgation of reading news, balance sheets and financial ratios. The stock market prediction using news or personal opinion has attracted much attention of academician and business.

The main challenge is the prediction of stock market values to extract meaningful characteristics from news articles and decide which stock to buy, hold and sell. The news articles generally contain both positive and negative aspects of stock market values, which make it difficult for financial analysts to predict the underlying truth. Such analysts have to deal with vast information available in newspapers, magazines and other resources in order to investigate the stock trend movement to make a prediction model.

Text mining technology has increasingly becoming important for easy and quick availability of news information which can be accessed due to sheer volume of opinion rich web resources. In recent years, our social and political system has a great impact of public sentiments and emotions. Similarly, the postings of social media have helped reshaping businesses. Text opinion mining is designed to help investors to predict the hidden patterns from the data available in digital form.

Opinion Mining and sentiment analysis are the techniques used to detect and extract subjective information from text documents. The sentiment classification is the main challenge in opinion mining which depends on mood, judgment or evaluation of any object like film, book etc. in the form of a text document, sentence or features which can be labeled as positive or negative.

We are conducting a study on predicting the long-term stock prices using Fundamentals, Text mining, Machine Learning. The need of the study is to come up with a one-stop shop for long-term stock prediction.

Our study is organized into following sections: Data collection used for generating historic data, fundamental analysis, sentiment analysis techniques, prediction algorithms used to predict the stock movement and the evaluation measures. Last section concludes the overall survey and discusses some future aspects of research.

## **Importance of Stock Price Prediction**

## **Important Terminologies in Stock Market**

## **Financial Ratios**

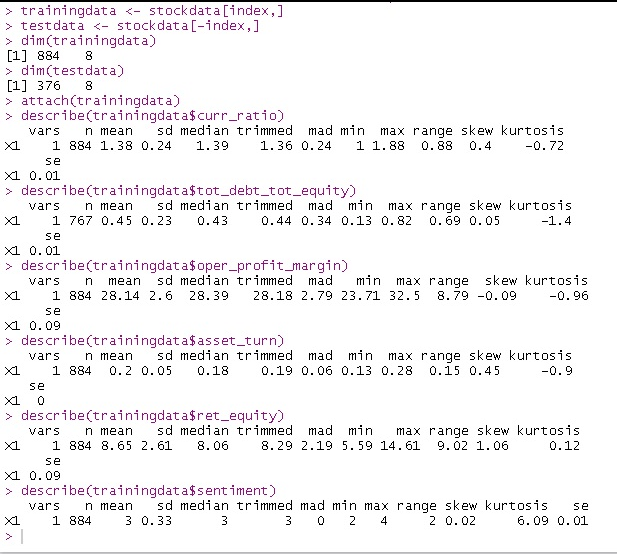
# **Chapter 3**

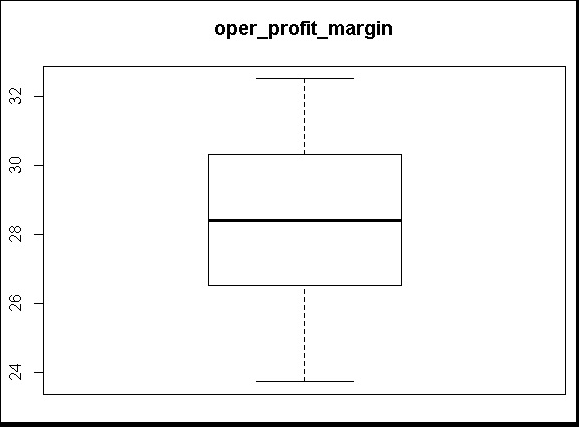
## **Data under Study**

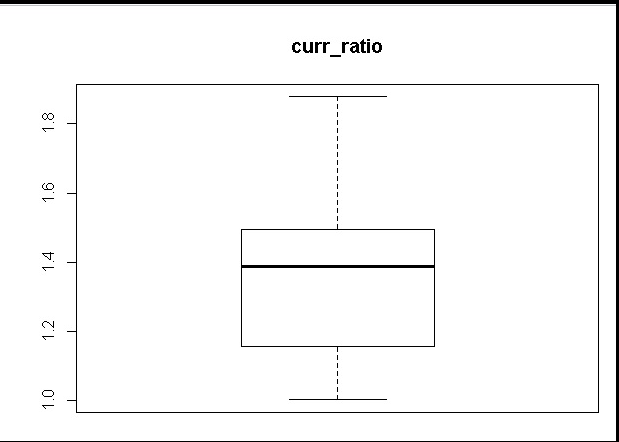
## **Analytical Approach**

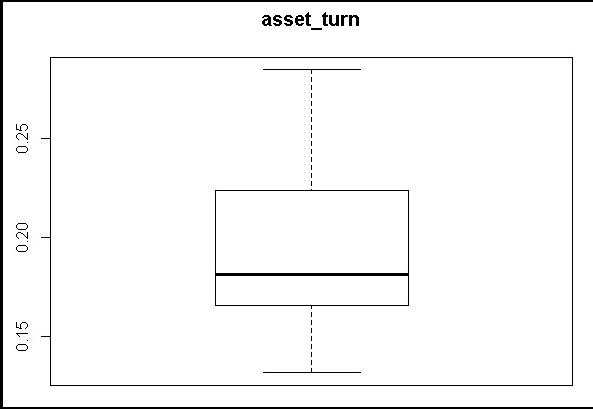
## **Exploratory Data Analysis**

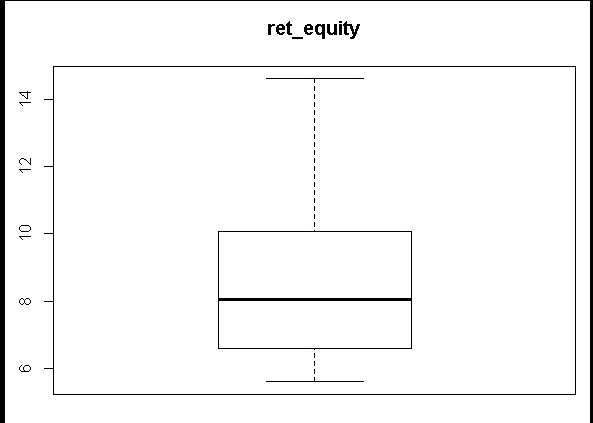
We have picked few sample stocks for EDA and here are the findings.

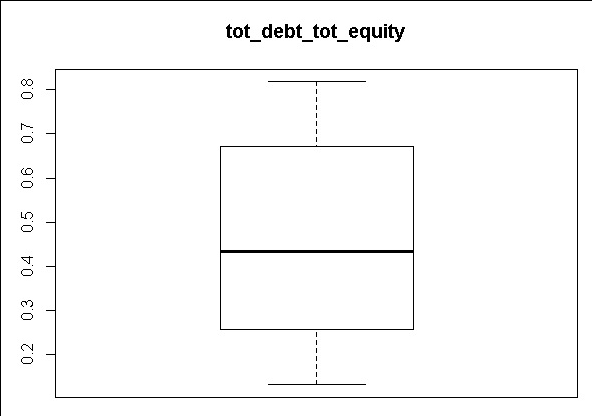


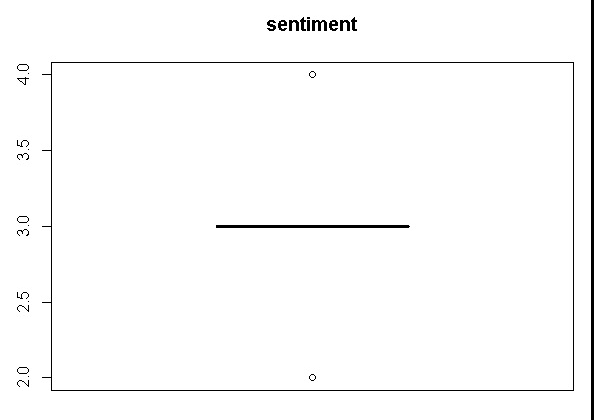


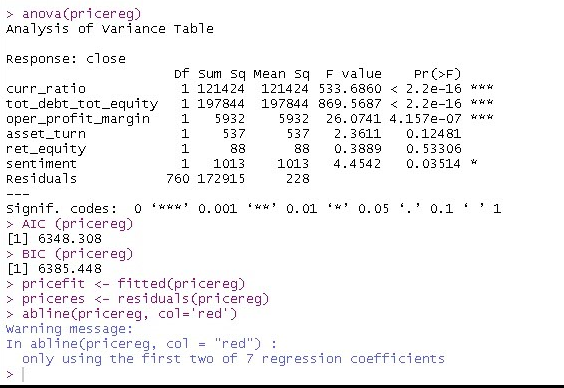












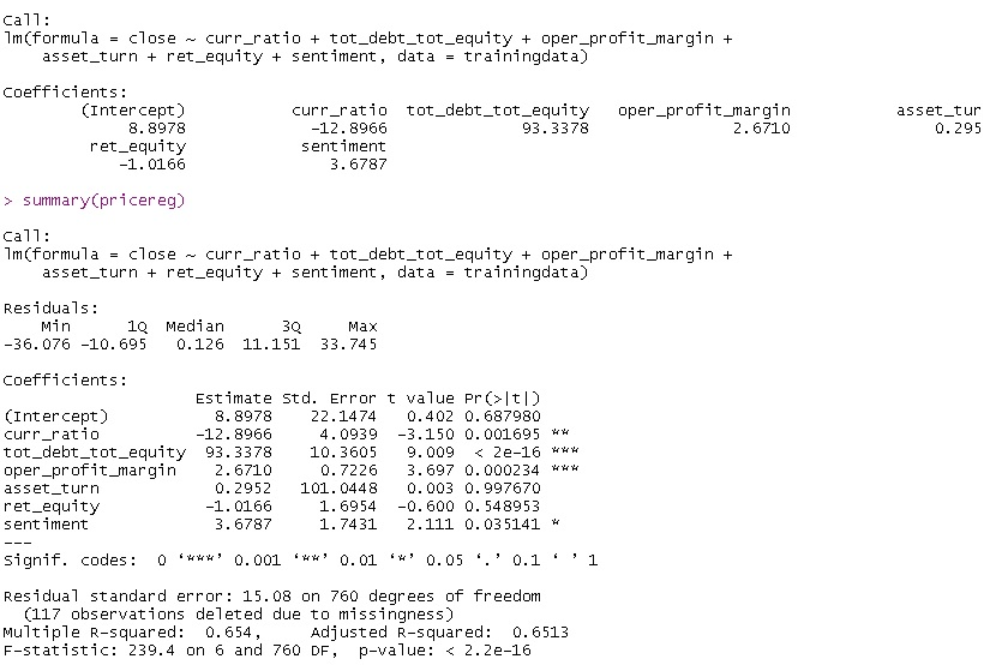
# **Chapter 4**

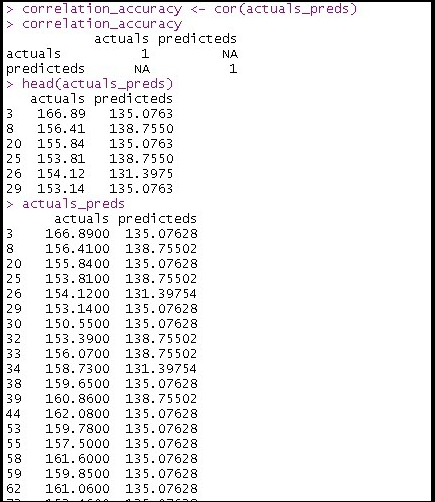
## **Sentiment Analysis**

## **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.

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

# **Chapter 5**

## **Recommendations**

## **Conclusions**

# **Bibliography**

# **Annexure**