

Financial Forecasting using Deep Learning

Yash, Shruti, Aishwarya, Mangesh and Sandeep Gore

*Department of Computer Engineering
Student at GHRCEM, Wagholi.
Pune, Maharashtra India, 412207*

*Department of Computer Engineering
Professor at GHRCEM, Wagholi
Pune, Maharashtra, India 412207*

Abstract

Investing in the stock market is complex and challenging for people because prices are changing every second. For that, they have to update with the latest financial news. But nowadays, understanding financial information or news is hard.

To solve this problem, this paper proposes a model of stock price forecasting based on financial news sentiment. The model is using Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Using the latest research in Natural Language Processing, the model analyzes the sentiment of financial news, and then autoregressive integrated moving average model (ARIMA) forecasts the result. For this model, data is coming from Twitter and Yahoo API which after preprocessing going through a model.

Key-words: Data analytics, Sentiment Analysis, News Articles ,LSTM-RNN, Arima, Deep-Learning.

1. INTRODUCTION

Stock price prediction is nowadays hot topic among the investor and youngster. Stock prediction is important topic in economics and finance. The stock market of the world is very fluctuating in nature. They fluctuate based on various factors as previous stock prices, financial news etc. Additionally, stock market also affects by non-economic factors such as natural disaster and political decision.

To predict the stock market investor uses either technical or fundamental analysis. In technical analysis we directly analyzing from the price and in fundamental analysis we analyzing by unstructured data like financial news, blogs, tweets, earning report.

With the newest technological developments and advancement in data analytics, financial professionals and economists have increasingly explored new AI and machine

learning approaches to reinforce financial market forecasting results.

In general machine learning may be a term used for all algorithm's methods using computers to reveal patterns based only on data and not using any programming instructions. The machine learning approach is can be sub divided into two sub-categories supervised learning and unsupervised learning algorithm. In supervised machine learning the program is "trained" on pre-defined set of coaching examples, which then facilitate its ability to succeed in an accurate conclusion when given new data. The training data is labelled with the answers. In Unsupervised machine learning, the program is given a bunch of data and must find patterns and relationships therein. A collection of unlabeled data is given, which is to be analysed and discover patterns within.

Deep learning is the part of machine learning in which "deep" word means that the number of levels or layer through the data is passed.

Deep neural networks are generally interpreted in the form of probabilistic inference. RNN is one of the deep learning techniques which is used in various applications in machine learning. LSTM is one of the supervised deep learning model which we are going to use for natural language processing..

2. RELATED WORK

Different dataset is used by different researchers to analysis the result through sentiment analysis. Some work is discussed in this section.

This paper [1] studies online movie reviews using sentiment analysis approaches. Sentiment classification techniques were applied to movie reviews. Specifically, we compared two supervised machine learning approaches SVM, Naïve Bayes for Sentiment Classification of Reviews.

This paper[2]studies the stock market assests to predict the future value of the google and nike stocks .using the lstm-rnn machine learning algorithm.and they got the conclusion.

This paper[3]studies architecture that can handle typically-long and content-specific news articles, which often cause overfitting when trained with neural networks. Moreover, the proposed architecture can also effectively process the case when the subject to be analyzed sentimentally is not the main topic of the concerned article, which is also a common issue when performing aspect-level sentiment processing.

This paper [4] aims to provide a comprehensive survey on financial sentiment analysis (FSA) including data source, lexicon-based approach, traditional machine learning approach and recent deep learning approach like word embedding, CNN, RNN, LSTM and attention mechanism. Our inspirations in future direction like large unsupervised contextual pre-training, hierarchical coarse-to-fine approach, joint learning, transfer learning and possible applications are also discussed.

This paper[5] they build an NLP system which process the free-text data and extract NGS-related information. They used three types of recurrent neural network (RNN) VIZ, gated recurrent unit, long short-term memory (LSTM), and bidirectional LSTM to classify documents to the treatment-change and no-treatment-change groups

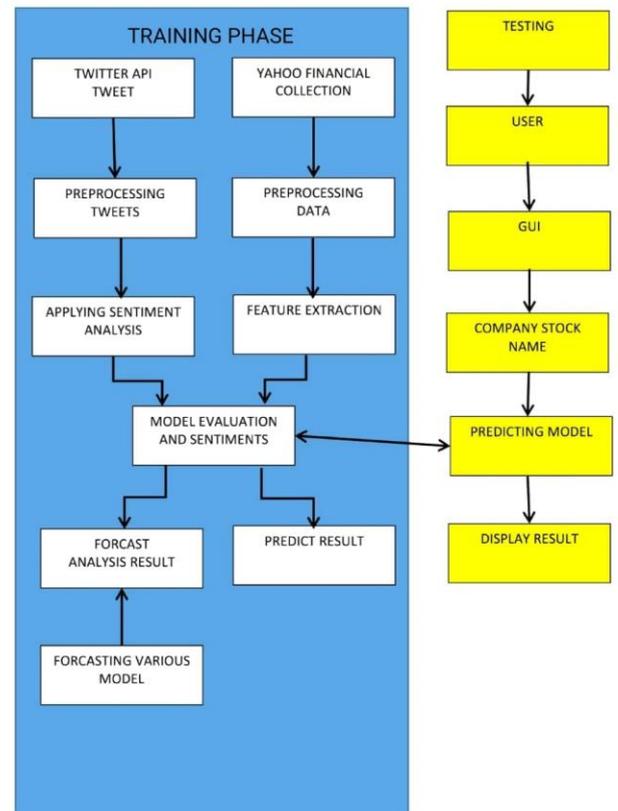
This paper [6] they showed that LSTM RNN architecture achieve state-of-the-art performance for large scale modeling. They proposed deep LSTM RNN architecture outperforms standard LSTM networks and DNNs and makes more effective use of the model parameters by addressing the computational efficiency needed for training large networks.

3.Proposed model:

This paper is based on the model which forecast the Stock price with the help of financial news articles

Flow for the proposed system is here in this diagram.

We will collect data from twitter and yahoo financial news and will forecast the result by processing and analyzing the data



4. METHODOLOGY

4.1 Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM)

Deep Learning is a subpart of machine learning .the machine learning can be supervised learning unsupervised learning. RNN is a deep learning architecture used for various application in this paper we are going to use it for Natural Language processing. It is the first algorithm which remembers its input which helps them to predict whats coming next.also known as feedforward model

In deep learning there are several levels which individually work on input data to make it more usable. In this we are taking raw data as financial news articles. RNN comes in various variant such as fully recurrent, LSTM, GRUs ,Bi-Directional.We are using LSTM here which helps us to remove vanishing gradient problems. Using LSTM units can train model in supervised fashion.

LSTM is consist of three layer one is input other is output and third one is hidden layer. that hidden layer contains various gates and cell units. The LSTM can read, write and delete information from its memory. In an LSTM you've three gates: input, forget and output gate. These gates determine whether or to not let new input in (input gate), delete the knowledge because it isn't important (forget gate), or let it impact the output at the current timestep (output gate). Below is an illustration of a RNN with its three gates: The gates in an LSTM are analog within the sort of sigmoid, meaning they vary from zero to at least

one . The fact that they're analog enables them to try to back propagation.

Before using the dataset for lstm model we have pre-processed it using different steps which are used traditionally .

4.2.ARIMA Model

ARIMA stands for 'Auto Regressive Integrated Moving Average'. this model is a widely used (statistical) forecasting method for time series prediction. So we fitted the model to the time series data for analyzing the information or to predict the future data points on the time scale. The time series forecasting model is capable of predicting future values supported by previously observed values.

There are three components that are AR, I, Ma are also standard input parameters.

"AR" means Auto-Regressive which uses the dependent relation between an Observation and a few predefined number of lagged observations also called delay .

"I" mean Integrated which subtracts an observation from observation at the previous time step so as to form the statistic stationary.

"MA" means Moving Average which exploits the connection between the residual error and the observations.

The general process of ARIMA model is the following:

1. Visualize the time series data
2. make the time series data stationary
3. plot the Correlation and AutoCorrelation Charts
4. Construct the ARIMA mode Based on data
5. use the model to make a prediction

5.EXPERIMENT AND RESULTS

5.1. Data Collection :

We are fetching the news from twitter by using the twitter API. These fetched tweets are contained in JSON format. For the stock price we are using NASDAQ and NSE stocks. Stock ticker is taken as input from user.

5.2.JSON to String Conversion:

The collected tweets from twitter are in the format of JSON and we need to convert it into the string. For that we are using full text method. Full text convert tweets into the string.

5.3.Data Cleaning:

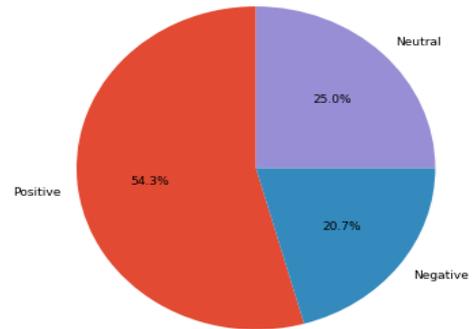
The text contain unnecessary words that may not be required for sentiment analysis. The text may contain punctuation and accents, special characters, numeric digits, whitespace, emojis, hashtags, hyperlink. To remove these unwanted list, the tweet-preprocessor has been used.

5.4. Text Blob library:

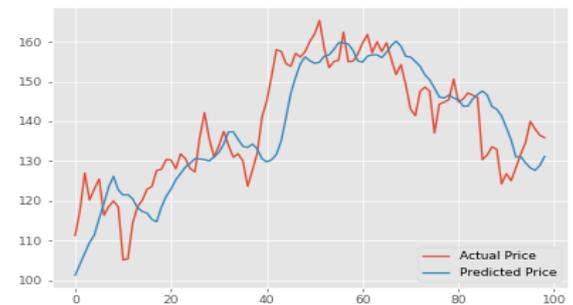
Text Blob is a python library for Natural Language Processing(NLP). TextBlob actively used Natural Language Toolkit(NLTK) This

library used for processing the textual data. Sentiment model helps to predict the polarity of text. Polarity are classify into positive, negative or neutral. TextBlob library is used to assign the polarity as positive, negative, neutral to processed tweets.

EXPERIMENTAL RESULTS AND ANALYSIS:



SENTIMENT ANALYSIS FOR CANBK.NS TWEETS



LSTM MODEL ACCURACY FOR CANBK.NS

LSTM model was applied to the test set data that is 20% of the entire dataset. The predicted values were compared against the actual values



RECENT TRENDS IN CANBK.NS STOCK PRICES

The NSE stock data for the past 2 years along with real time prices is fetched from the Yahoo Finance API.

Comparison of Model:

The Root Mean Squared Error (RMSE) of LSTM, ARIMA and LINEAR REGRESSION models for Canara Bank (CANBK.NS) stocks are compared below. It is seen that for LSTM and ARIMA model have low RMSE value compared to linear Regression model.

ALGORITHM	RSME
LSTM	8.78
ARIMA	5.18
LINEAR REGRESSION	11.46

6.FUTURE WORK:

The number of NLP papers being published each year is any indication, it would seem that NLP is a growing field .we have created model and are planning to implement model on different platforms .Also we can use different models to increase the accuracy of prediction

7.CONCLUSION

In proposed model we compare LSTM, ARIMA and Linear Regression algorithm for different NASDAQ and NSE stock and from above comparison, graphs and plots, it is observed that LSTM and ARIMA model gives more constant results and Linear Regression. Overall sentiment analysis of tweet gives a basic idea about the current trend in market. Fundamental analysis (news, unstructured text) and technical analysis (stock price) both are affects on market trend. Tweet and stock price both combined give recommendation whether to invest in particular stock or not.

References

- [1] Sentiment Classification using Machine Learning Techniques Suchita V Wawre¹, Sachin N Deshmukh² International Journal of Science and Research (IJSR) Volume 5 Issue 4, April 2016, 819 - 821
- [2] Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms. Suvarna G, Karuna C, Ashok K, P S Hiremath **Published in: 2020 International Conference on Inventive Computation Technologies (ICICT)**
- [3] Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm. Shuhaida Mohamed Shuhidan¹, Saidatul Rahah Hamidi ², Soheil Kazemian³, Shamila Mohamed Shuhidan⁴ and Maizatul Akmar Ismail⁵
- [4] Financial Sentiment Analysis(FSA): A Survey. Xiliu Man, Tong Luo, Jianwu Lin
- [5] Sentiment analysis algorithms and applications: A survey. WalaMedhat^aAhmedHassan^bHodaKorashy^b
- [6] Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters JitendraKumar^a RimshaGoomer^b Ashutosh KumarSingh^a
- [7] Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling Has,im Sak, Andrew Senior, Françoise Beaufays
- [8] M. Wollmer, F. Eyben, B. Schuller, and G. Rigoll, "A multi-stream asr framework for blstm modeling of conversational speech," in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, May 2011, pp. 4860–4863.