Stock Market Prediction And Recommendation System

Review Paper

Prof. Aarti Puthran, Ashutosh Shukla, Devansh Shah, Khushboo Dattani.

[1] Teacher, Student,Student,Student.
[1] Information Technology
[1] Shree L.R. Tiwari College of Engineering, Mumbai, India

Abstract: Stock Market's nature is very Dynamic. Even the smallest news or tweet can affect the prices of shares. The prices of shares depend on various factors. Investing in the stock market comes with a risk. Hence, the idea to get a prediction model, with the intention of people being more confident while investing their money comes from here. To develop a model that would help to predict stock prices, we would have to look at multiple things like collecting the data, cleaning the data, understanding algorithms, looking out for best methods to predict. In this paper, we are mainly comparing the algorithms that we selected for building a model on the basis of various parameters, also here we are mentioning the factors on which the analysis can be done and further we discuss the fuzzy model concept which would help combine the results of two algorithms.

Index Terms - Prediction, Analysis, Fuzzy model.

I. INTRODUCTION

From the beginning, the Stock Market has attracted many people. As it was giving the fastest income to the people. But due to the high risk factor, majorly the richer and upper middle class peoples were taking part in it.

The first lockdown in April 2020, has affected lives of many people and also they lost their source of income. This situation forced the poor people to risk their money in the share market in order to make money for their livelihood. Nowadays, many people have been involved in Share Market and it has become an attractive topic to talk about.

This project makes use of different algorithms and makes use of previous stock prices and news from google finance to predict whether the particular share should be bought or sold.

3.1 Data and Sources of Data

All the data regarding stock prices i.e opening price, closing price, high price and low price is taken from MoneyControl website. All the news headlines for the stock are taken from Google Finance. 4 previous headlines will be considered for a particular stock.

3.2 LITERATURE REVIEW:

1. Stock Prediction using Sentiment analysis and Long Short Term Memory-[1]
In this paper, the author combines machine learning and public sentiment. This has been achieved by using a hybrid algorithm that uses sentiment analysis and LSTM to predict the next day stock values and the public's sentiment, which helps us to correlate the market conditions and public sentiment. Publicly available Twitter data is used to perform sentiment analysis and yahoo finance to get stock values.

2. Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis- [2]
In this paper the author suggests the use of a comprehensive method to predict the state of stock market over traditional approach.This because of the more accurate prediction achieved by this. Author suggests analysing both the historical data and current sentiment regarding the stock to recommend buying or selling the stock. This is feasible because of the availability of great computational powers and lots of data. Author divides the whole system into three modules:
1.Machine learning module for analysing the historical data
2. Sentiment Analysis module for analysing the news
3. Fuzzy logic module for combining the results of both and giving recommendations based on that.

3. Stock Price Prediction Using News Sentiment Analysis- [3]

In this paper analysing the stock market based only on historical data is deemed as insufficient by the author. The author says so because stock prices can deviate greatly due to political events, government's economic policies, news etc which cannot be tracked by the data analysis alone. Hence the author suggests use of sentiment analysis to predict whether the stock price will go up or down. Sentiment analysis is to be done by using various news sources such as twitter data,financial news websites, etc as a source of raw data.

4. Recent Advances in Stock Market Prediction Using Text Mining: A Survey- [4]
In this paper a review of background concepts such as Sentiment analysis, twitter sentiment,online financial news sentiment, textual data preprocessing, feature extraction out of it, feature representation. Also, a relationship between stock market prediction and text mining. This paper reflects that there is a strong relationship between market movements and information published in news and social media.

5. Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty-[5]
This paper brought into light different Sentiment Classification Methodologies, Machine Learning Approach which is further classified into supervised and unsupervised learning, Lexicon based approach which is further divided into dictionary based approach and corpus based approach and so on. It showed sentiment analysis for the stock market by fetching Sensex and Nifty live server data values on different intervals of time that can be used for predicting the stock market status. For this purpose they use a python scripting language which has a fast execution environment.

3.3 Fuzzy Model

The Fuzzy Module is a modern approach towards stock market data analysis which uses both stock market prices and general sentiment of a stock in the market as an input. It combines the results of the both to provide a more accurate and reliable recommendation. This method has been proven to be better than the traditional approach to decide whether a stock should be bought or sold

The purpose of this module is to output Stock Faith which is the strength of recommendation. The activation rules for this module are as follows:
• IF the News Sentiment was good or the Stock Prediction value was good, THEN the Stock Faith will be high.
• IF the Stock Prediction value was average, THEN the Stock Faith will be medium.
• IF the News Sentiment was poor and the Stock Prediction value was poor, THEN the Stock Faith will be low.

3.4 Comparison Models for Data Analysis

Parameters	LSTM (Basic rate Prediction)	CNN
Complexity	High	Low as compared to LSTM
Used for	Prediction of Sequential data	Mostly Used For Images, Speech

Long Short Term Memory (LSTM):
It is a type of recurrent neural network that is capable of learning order dependencies in sequence prediction problems. It has feedback connections which help it to process sequences of data. It also has internal state cells which function as long or short term memory cells. The output is regulated by these state cells. This property is very useful when we need to depend on prior inputs rather than the latest ones. As time passes it is less likely that the output will be dependent on very old inputs, these inputs are forgotten by it through their forget gates, the forget gates is just a multiplicative factor of 0.9, that is within 12 steps the factor becomes $0.9^{12}= 0.282$. Equations used by forget, input and output gates in LSTM are given below. These equations help LSTM to consider previous inputs in predicting the next output and also the forget gate equation helps in forgetting very old inputs.

$$f_t = \sigma(w_f (h_{t-1}, x_t) + b_f)$$
$$i_t = \sigma(w_i (h_{t-1}, x_t) + b_i)$$
$$o_t = \sigma(w_o (h_{t-1}, x_t) + b_0)$$

The data is then split into a training and testing data set. After converting and storing the data as numpy arrays it is then trained under the LSTM model. We used Adam as our optimizer and trained 4 epochs. The predicted close values were then compared to the actual close values and Root Mean Squared Error was calculated in order to determine our accuracy and a chart is prepared in order to visualize the comparison between the predicted and actual values

Convolutional Neural Network (CNN):

Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, a rough imitation of their biological counterparts, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value.The behavior of each neuron is defined by its weights. When fed with the pixel values, the artificial neurons of a CNN pick out various visual features.

When you input an image into a ConvNet, each of its layers generates several activation maps. Activation maps highlight the relevant features of the image. Each of the neurons takes a patch of pixels as input, multiplies their color values by its weights, sums them up, and runs them through the activation function.

The first (or bottom) layer of the CNN usually detects basic features such as horizontal, vertical, and diagonal edges. The output of the first layer is fed as input of the next layer, which extracts more complex features, such as corners and combinations of edges. As you move deeper into the convolutional neural network, the layers start detecting higher-level features such as objects, faces, and more.

3.5 Comparison Models for Sentiment Analysis:

Parameters	Naive Bayesian (ML Approach) (Sentiment Analysis)	Lexicon Method
Accuracy	87%when used with LSTM	50%
Complexity	Low fast computational power	High
Used for	Text Classification And With Problems Having Multiple Classes.	Finding Extreme Opinions

Naive Bayes classifier:

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.
Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Lexicon Method:

It involves calculating the sentiment from the semantic orientation of word or phrases that occur in a text [25]. With this approach a dictionary of positive and negative words is required, with a positive or negative sentiment value assigned to each of the words. Different approaches to creating dictionaries have been proposed, including manual [26] and automatic [27] approaches. Generally speaking, in lexicon-based approaches a piece of text message is represented as a bag of words. Following this representation of the message, sentiment values from the dictionary are assigned to all positive and negative words or phrases within the message. A combining function, such as sum or average, is applied in order to make the final prediction regarding the overall sentiment for the message.
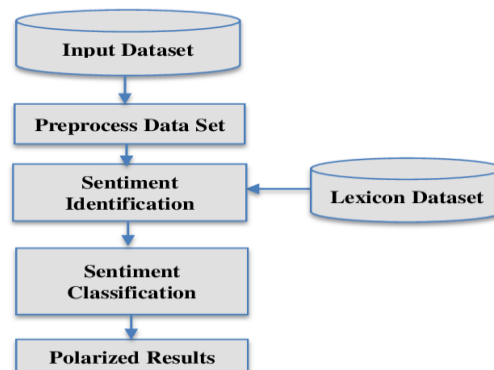
IV. CONCLUSION

In conclusion a system is made dedicated to predicting the future price of a particular stock and recommending whether a stock should be bought or sold. These predictions are based on previous prices and recent news headlines about a particular stock. Having such a smart system would result in a more risk-free investment. Due to this system people will perceive investing in the stock market as an opportunity rather than a risk which is the ultimate goal of the project.

Even the people who are ignorant of the way the stock market works can easily take guidance from this system before investing. This makes it useful for everyone.

REFERENCES

[1] Harsh Panday, V. Vijayarajan, Anand Mahendran, A. Krishnamoorthy, V.B. Surya Prasath, "Stock Prediction using Sentiment analysis and Long Short Term Memory", European Journal of Molecular & Clinical Medicine, 7-February 2020.

[2] Ashish Pathak, Nisha P shetty, "Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis", Conference paper, 4th July 2018.

[3] Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, David C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis", 4-9 April 2019.

[4] Faten Subhi Alzazah, "Recent Advances in Stock Market Prediction Using Text Mining: A Survey", June 1st 2020.

[5] Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawan, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", 2015