WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Price Prediction using Time Series Analysis and Sentimental Analysis

Suyash Jain, Yagvendra Khichi, Yashwant Patidar, Yogesh Sharma

AITR,Indore
major.projectstonks@gmail.com

ABSTRACT

The prediction of the stock price is an important and thriving topic in educational research and financial engineering. It has always drawn the attention of people across the world having an interest in earning financial or monetary blessings through investment in the share market. Stock market forecasting is a difficult task that is influenced by a variety of elements such as corporate financial status, national legislation, and so on. Stock prices are affected by a variety of factors these days, including company-related news, public perception, public sentiment, and social-economic, and political conditions in the nation. This software is designed to eliminate or, in some circumstances, mitigate the difficulties that these systems now encounter to an extent. Furthermore, this software is tailored to the people's specific requirements for smooth and efficient operations.

INTRODUCTION

There are numerous sides to the stock prediction problem. Price prediction, according to the Efficient Market Hypothesis (EMH), are mostly driven by fresh information and follow a random walk pattern. Several people have tried to extract changes or patterns in the way stock markets operate and react to various stimuli, despite the fact that this theory is largely acknowledged by the scientific community as a basic paradigm regulating markets in general.Researchers having keen interest in investment opportunities finds it extremely intriguing to work on. People's sentiment, perception, corporate performance, news, and regulators all influence the pricing value. Machine learning-based stock price prediction assists in determining the future worth of a company's stock and other financial assets traded on an exchange.

Stock market forecasting is a difficult undertaking since it necessitates in-depth knowledge of events, historical data analysis, and the impact of news on stock price movements. The problem is made even more difficult by the extreme volatility of stock price patterns.

The whole point of stock price forecasting is to earn money with a calculated assessment of risk.However, a comprehensive analysis of the general background of stock prediction is lacking in the literature.

The tool "Stonks" was created to address the issues that arise in the practice of stock prediction. In the current machine learning arena, stock price prediction is an important task. To tackle this challenge, several ways have been presented. The majority of them are focused on a stock's time series analysis, while some are based on news feelings to predict a stock's value. Our tool provides both options to the user i.e., prediction as well as sentiment about the company on Twitter

PROBLEM FORMULATION

We decided to embark on this project for the following reasons:

1. Existing systems employ a variety of techniques, ranging from regression to classification. The existing approach fails when there are rare outcomes or predictors because the algorithm is based on bootstrap sampling, which has a number of limitations.
2. Previous results show that when the standard classifier is applied, the stock price is uncertain.
3. Existing systems either consider historical share price data or sentimental analysis for prediction but does not provide a single solution based on both services.

LITERATURE REVIEW

In the topic of stock price prediction, a lot of effort has been done; yet, getting the desired outcomes is not straightforward.Following are some of the significant conclusions drawn from the literature review:

Stock market forecasting is a difficult undertaking since it necessitates in-depth knowledge of events, historical data analysis, and the impact of news on stock price movements. The problem is made even more difficult by the extreme volatility of stock price patterns.

Several studies based on stock price prediction also goes into the usage of approaches for extracting opinions. It also highlights the application of domain knowledge in both textual feature extraction methodologies. It also emphasises the need of using deep neural network-based prediction approaches to uncover the hidden relationship between textual and numerical data. These studies are significant and unique in that it develops a comprehensive framework for stock market forecasting and identifies the strengths and flaws of existing methods. It covers a wide range of open issues and research directions that the research community may find useful.

There are several studies such as Forecasting Stock trend using Time Series data, using Textual Data, using Sentimental Analysis, Using Numerical and Textual Data. Each has different methods for prediction and considers various parameters and precision.

Stock market patterns are exceedingly volatile, which makes forecasting difficult.

METHODOLOGY

1)Time Series Analysis (TSA)

Time series analysis is a statistical method that analyses past data within a specific period to predict the future. It consists of an ordered sequence of evenly spaced data.

Let's look at an example to understand time series data and analysis. Consider an example of passenger information. There is a certain number of passengers for a certain period.

Timeseries forecasting models are the models that are capable to predict future values based on previously observed values. Timeseries forecasting is widely used for nonstationary data.

Nonstationary data are called the data whose statistical properties e.g., the mean and standard deviation are not constant over time but instead, these metrics vary over time.

These unsteady input data (used as inputs to these models) are commonly referred to as time series. Examples of time series include
temperature values over time, stock prices over time, and house prices over time. Therefore, the input is a signal (time series) defined by time series observations.

2) Long Short-Term Memory (LSTM)

LSTM networks are extensions of recurrent neural networks (RNNs) introduced primarily to handle situations where RNNs fail. An RNN is a network that processes the current input by taking into account the previous output (feedback)
and storing it in memory for a short time (short-term memory). Of the various applications, the most popular are in the areas of speech processing, non-Markov control, and music composition. Still, RNNs have their drawbacks.

First, the information cannot be stored for long periods of time. You may need to refer to certain information stored long ago to predict current performance. However, RNNs cannot fully handle such "long-term dependencies".

Second, there is no finer control over what part of the context needs to be continued and how much of the past needs to be "forgotten". Another problem with RNNs is the explosion and disappearance of gradients that occur during
the backtracking network training process (discussed later). This is how Long Short-term Memory (LSTM) came into being. It is designed to almost completelyeliminate the vanishing gradient problem without changing the training model.

Long delays in certain problems are filled with LSTMs. In this case, the LSTM also handles noise, distributed representation, and continuous values. LSTMs do not have to hold the finite number of states required by Hidden Markov Models (HMMs).

LSTMs provide a wide range of parameters such as learning rate and input / output bias.

Therefore, no tweaks are needed. The complexity of updating each weight is reduced to O (1) using LSTMs. This has the same benefits as Back Propagation Through Time (BPTT).

3) Sentimental Analysis

Sentiment analysis is the process of classifying whether a block of text is positive, negative, or neutral. Sentiment analysis is a word context mining that shows a brand's social sentiment and helps determine if the products it manufactures are in demand in the market.

The goal that sentiment analysis is trying to achieve is to analyse people's opinions in ways that help the company grow. It focuses not only on polarity (positive, negative, neutral) but also on emotions (happiness, sadness, anger, etc.).

It uses a variety of natural language processing algorithms, including rule-based, automatic, and hybrid

Rule-based sentiment analysis refers to a survey conducted by a language expert. The result of this study is a set of rules (also known as lexicons or mood lexicons) that the
classified word is either positive or negative, and the corresponding intensity measurements.

Generally, the following steps are needed to be performed while applying the rule-based approach:

1.Extract the data

2.Tokenize text. The task of splitting the text into individual words

3.Stop words removal. Those words which do not carry any significant meaning and should not be used for the analysis activity. Examples of stop words are: a, an, the, they, while etc.

4.Punctuation removal (in some cases)

5.Running the pre-processed text against the sentiment lexicon which should provide the number/measurement corresponding to the inferred emotion.

4) Flair

Flair is a lightweight, open-sourced natural language processing (NLP) library developed by Zalando Research. The Flair framework is built directly on top of one of the best deep learning frameworks, PyTorch.

The Zalando Research team has also published some pre-trained models for the next NLP task.

1.NameEntity Recognition (NER): Can recognize whether a word represents a person, place, or name in the text.

2. Part-of-speech tagging (PoS): Marks all words in the specified text according to the "part of speech" to which they belong.

3. Text classification: Classification of text based on criteria (labels)

4.Custom model training: Build your own custom model.

The Flair library contains many great features. This is my choice of the most prominent one:

1.Includes popular, state-of-the-art word embeddings such as GloVe, BERT, ELMo, and character embedding. Very easy to use thanks to the Flair API.

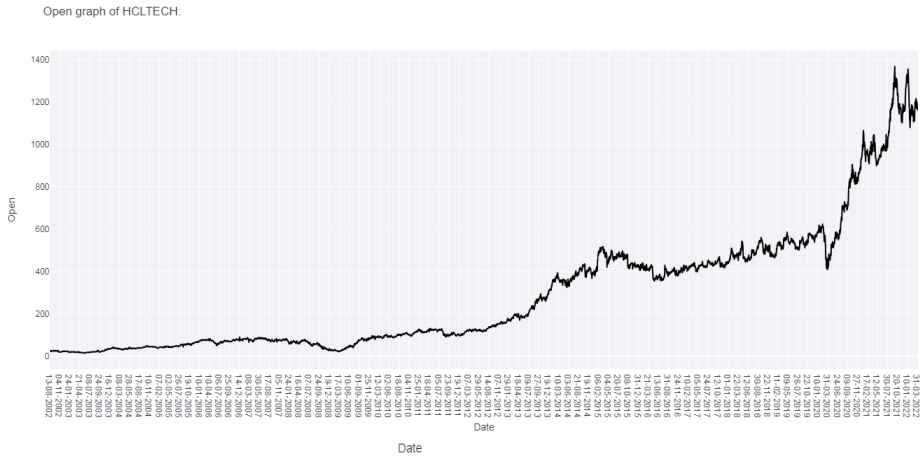
2.Flair's interface allows you to combine different word embeddings and use them to embed documents. This will greatly increase the results.

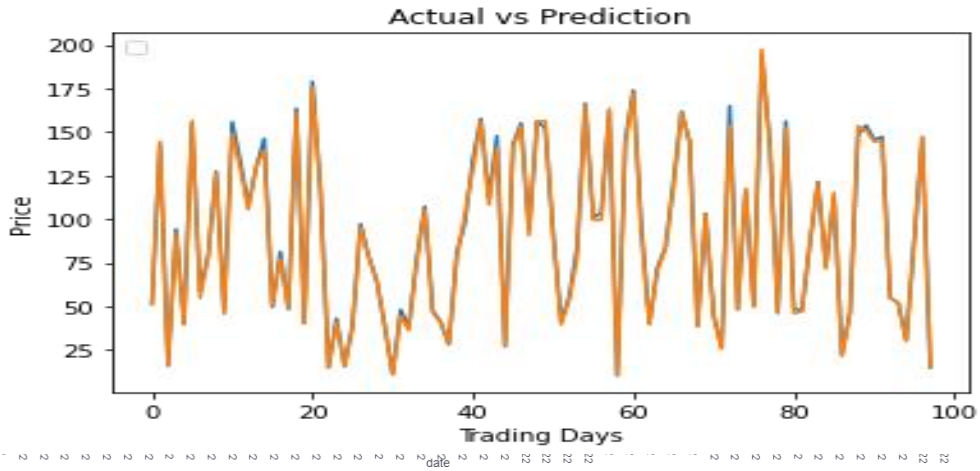
3."Flare embedding" is a signature embedding provided by the flare library. This is supported by embedding the context string.

4.Flair supports a variety of languages-and is always looking for new languages

RESULT DISCUSSION

The tool “Stonks” will assist investors in making stock market investments based on a variety of parameters. The project's goal is to develop an application that can analyse a company's previous year's data and input these values into a model that can predict the value of a stock in the near future with reasonable accuracy and should also be able to tell or give visual information to user about the public perception i.e., public sentiments about the company on the day of using its services so as to take final decision before investing considering both, the data-based prediction and public sentiment.





CONCLUSION

We suggested a more accurate model that employs RNN and LSTM to forecast the trend in stock prices and uses sentimental analysis to extract the perception of people about a company. In the buried layer of the network, LSTM introduces the memory cell, a computational unit that substitutes typical artificial neurons. The accuracy of prediction is improved in this work by increasing the Epochs and batch size. We use test data to forecast in the proposed method, and the predicted outcomes are more accurate than the test data. The proposed method can track and anticipate the stock market, with the prediction yielding greater and more accurate outcomes. We are getting accurate findings in our above model, which will be more valuable to stock analysts, business analysts, and other professionals.

ACKNOWLEDGEMENT

Without mentioning the people who made it possible, whose continual direction and support crowned our efforts with success, the happiness and exhilaration that accompanies the successful completion of any endeavour would be impossible. We would like to show our gratitude to the Professor Praveen Bhanodia, Department of Computer Science and Engineering, our guide, for his support, encouragement and wisdom shared with us during the course of development of our project and this research Finally, we'd like to thank our institution for providing us with this opportunity to learn even more outside of our curriculum.

REFERENCES

[1].Shriram.S, Dr. K. Anuradha, Dr. K.P. Uma. “Future Stock Price using Recurrent Neural Network, LSTM and Machine Learning”, 2021

[2].Usmani S, Shamsi JA. “News sensitive stock market prediction: literature review and suggestions”, 2021

[3]. Dev Shah, Haruna Isah and Farhana Zulkerine. “Stock market Analysis: A Review and Taxonomy of Prediction Techniques”, 2019

[4]. K. Hiba Sadia, Aditya Sharma, Adarrsh Paul, SarmishthaPadhi, Saurav Sanyal. “Stock Market Prediction Using Machine Learning Algorithms”, 2019.

[5]. Ashish Sharma, Dinesh Bhuriya, Upendra Singh. “Survey of Stock Market Prediction Using Machine Learning Approach”, 2017

[6].Kordonis, John &Symeonidis, Symeon&Arampatzis, Avi. ” Stock Price Forecasting via Sentiment Analysis on Twitter”, 2016

[7]. Kalyani Joshi, Bharthi H.N, Jyothi Rao. “Stock Trend Prediction using News Sentiment Analysis”, 2016

[8]. Anshul Mittal, Arpit Goel. “Stock Prediction using Twitter Sentimental Analysis”, 2011