TITLE OF PROJECT
STOCK PRICE PREDICTION USING ML

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

CLOUD COMPUTING

Submitted by:

AYUSHMAN SINGH
20BCS4154

PARAG SHARMA
20BCS4158

KUNAL SINGH
20BCS4157

Under the Supervision of:

Dr. U. HARIHARAN

CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

APEX INSTITUE OF TECHNOLOGY

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

May & 2022

DECLARATION

I, **AYUSHMAN SINGH**, **PARAG SAHRMA, KUNAL SINGH**, student of **'Bachelor of Engineering in Computer Science'**, **session: 2022-23** ________, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled '**STOCK PRICE PREDICTION USING MACHINE LEARNING'** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

(Candidate Name) Candidate UID:

(AYUSHMAN SINGH)20BCS4154

(PARAG SHARMA) 20BCS4158

(KUNAL SINGH) 20BCS4157

Date:15 May 2022

Place:Gharuan

Annexure-3 (A typical specimen of table of contents)

Table of Contents

	Title Page	i
	Declaration of the	ii
	Student Abstract	iii
	Acknowledgement	iv
	List of Figures	v
	Timeline / Gantt	vi
	Chart	vii
1.	**INTRODUCTION***	
	1.1 Problem Definition	1
	1.2 Project Overview/Specifications* (page-1	2
	and 3)	3
		4
	1.3 Hardware Specification	4
	1.4 Software Specification	
	1.3.1	
	1.3.2	
	…	
2.	**LITERATURE SURVEY**	**8**
	2.1 Existing System	
	2.2 Proposed System	
	2.3 Feasibility Study* (page-4)	
		16
3.	**PROBLEM FORMULATION**	

4. **OBJECTIVES** 32
5. **METHODOLOGY** 45
6. **CONCLUSIONS AND DISCUSSION** 68
7. **REFERENCES** 85

ABSTRACT

In today's financial world stock exchange has become one of the most significant events. The world's economy today is widely dependent on the stock market prices. The Stock Market has been very successful in attracting people from various backgrounds be it educational or business .The nonlinear nature of the Stock Market has made its research one of the most trending and crucial topics all around the world.. People decide to invest in the stock market on the basis of some prior research knowledge or some prediction. In terms of prediction people often look for tools or methods that would minimize their risks and maximize their profits and hence the stock price prediction takes on an influential role in the ever challenging stock market business. Adopting traditional methodologies such as fundamental and technical analysis doesn't seem to ensure the consistency and accuracy in the prediction. As a result the machine learning technologies have become the recent trend in the stock market prediction whose prediction is based on the existing stock market values eventually as an outcome of training on their previous values. This paper focuses on RNN (Recurrent Neural Networks) and LSTM (Long Short term memory) technologies in predicting the ongoing trend of the stock market.

Keywords— **Stock, Stock Market, Shares, Shareholder, Recurrent Neural Network(RNN), Long Short Term Memory(LSTM)**

List of Figures

Figure Title

1.1 google stock graph

2.1 apple stock graph

3.1 IBM stock graph

4.1 tesla stock graph

5.1 LMS input and output

5.2 LMS updating weights

5.3 LMS updating weights

5.4 LSTM architecture

5.5 pre-processing of data

5.6 overall data

6.1 stock graph

6.2 stock graph comparison

6.3 predicted graph

6.4 predicted graph

6.5 stock graph

6.6 stock graph comparison

6.7 predicted graph

6.8 predicted graph

1INTRODUCTION

A stock or a share which is also known as a company's equity is referred to as a financial instrument that is used to represent an ownership in a company that represents a proportional assertion on its assets and earnings . Stock ownership means that the shareholder is the owner of a part of the company which is equal to the number of shares that is held as a fraction of the company's total outstanding shares. For example, an individual who is the owner of a hundred thousand shares of a company with a million outstanding shares would be having a ten percent stake ownership in it. The outstanding shares of most companies run into huge values as huge as millions or even billions.

Stock exchanges are nothing but the secondary markets wherein the present owners of the shares could transact with the potential buyers. It is a matter of utmost importance to understand that the corporations that have been listed on stock markets do not sell and buy their own shares often. So when a share of stock is bought in the share market, it is not being bought from the company itself but from the company's shareholder. In the same way when a share is being sold it is not being sold to a company directly but is being sold to an investor. In developing countries like India the rapid growth of its economy depends largely on the growth of its Stock Market. If there is a rise in the stock market, the growth in the company's economy would be rather high. If there is a downfall in the stock market the growth in the company's economy would be down. So it can easily be said that the stock market and country's economic growth is largely confined to the performance of the stock market. Only 10% of any country's population is showing interest or involving themselves with the stock market investment and the reason for this could majorly lie in the fact that the nature of the stock market is very dynamic. There is a misunderstanding or a false conception about the stock market and is often thought of as an act of gambling. So this misconception could be replaced by creating an awareness among people. Recently, many traders have been employing the machine learning models for the stock price prediction because of its rising popularity as a result of its proficiency and effectiveness in the prediction. Various methodologies such as sentimental analysis, usage of past prices of the stocks, growth in the sales and dividends have been developed to predict the stock effectively. As we are aware in order to predict the stocks accurately not only do we require data but also one of the above factors where the market hypothesis can be built. We have selected the Machine learning techniques here because they provide us with better results as compared to any other prediction model. It can be thought of as a proficient way to be able to represent such developments. It foretells a market value that is as close as possible to the perceptible value, hence improving its accuracy.

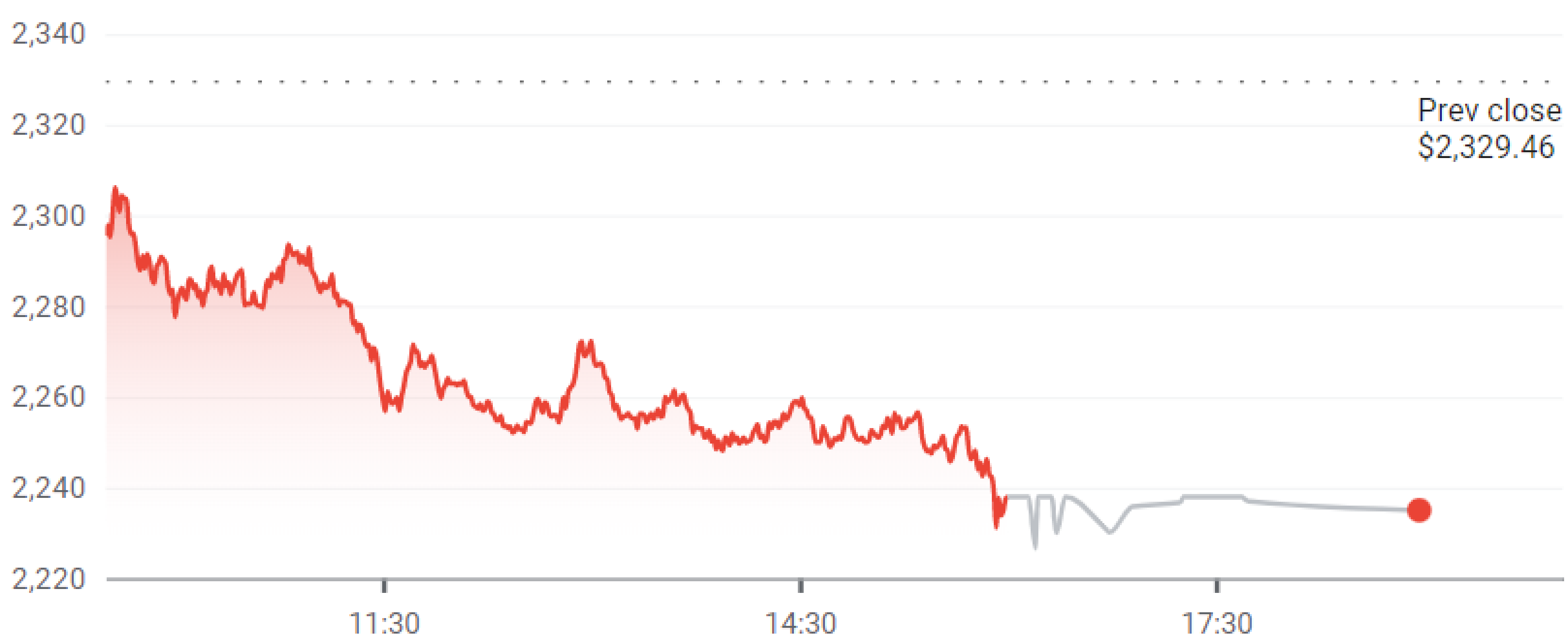

Fig. 1.1: google stock graph

The most important part of machine learning is the dataset that is used. The dataset being used must be as tangible as possible because even a slight variation in the data can cause considerable changes in the outcome. This paper majorly deals with predicting the stock prices using LSTM and RNN algorithms.

2 LITERATURE REVIEW

[1] Research on Stock Price Prediction Method Based on Convolutional Neural Network, IEEE 2019- Sayavong Lounnapha et al. This paper intends for a prediction model for stock price which is centered at the convolutional neural networks, that has exceptional capability of learning on its own. The data set is taught and tested relating the behaviours of both Convolutional Neural Networks and US stock market The result shows that the model on grounds of Convolutional

Neural Networks can effectually recognize the altering trend in stock market price and envisage it which provides significant allusion for stock price forecast. The accuracy of the prediction is found to be elevated, and it could also be promoted in the field of finance.

[2] Enhancing Profit by Predicting Stock Prices using Deep Neural Networks, IEEE 2019-Soheila Abrishami, et al., The prediction of economic time series is quite a herculean task, which has fascinated the attentiveness of many scholars and is extremely vital for investors. This paper focuses on presenting a deep learning system, which makes use of a range of facts for a part of the stocks to predict the value of the stock. This model has been trained on the smallest of data for a particular stock and accurately estimates the concluding value of that stock for multi-stepahead. It consists of an auto encoder in order to remove noise and makes use of time series data engineering to syndicate the advanced features with the original features. These new features are given to a Stacked LSTM Autoencoder for multistep-ahead estimation of the stock concluding value. Further, this estimation is used by a profit maximization approach to offer assistance on the right time for buying and selling a particular stock. The results indicate that the suggested framework

outclasses the state of the art time series forecasting methodologies with respect to analytical accuracy and effectiveness.

[3] Lately stock market has been the talk of the town with more and more people from academics and business showing interest in it. This paper mostly deals with the approach towards predicting stock prices using RNN (Recurrent Neural Network) and LSTM (Long Short Term Memory) on National Stock Exchange using numerous elements such as the present-day market price as well as anonymous events. A recommendation system along with models constructed on RNN and LSTM methods are used in selecting the company is also mentioned in this paper.

[4] The Stock Market Prices play a crucial role in today' economy. Researchers have discovered that social media platforms such as twitter and web news tend to influence the decisionmaking process of any individual. In this research behavioural reflex towards web news is taken into count to reduce the gap and make the prediction much more accurate. Precise

predictions were made for a day, a week and two weeks here after.

Fig. 2.1: apple stock graph

[5] The research work done by Mariam Moukalled Wassim El-Hajj Mohamad Jaber Computer Science Department American University of Beirut. Traditionally and in order to predict market movement, investors used to analyse the stock prices and stock indicators in addition to the news related to these stocks. Hence, the importance of news on the stock price movement. Most of the previous work in this industry focused on either classifying the released market news as (positive,

negative, neutral) and demonstrating their effect on the stock price or focused on the historical price movement and predicted their future movement. In this work, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news' sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted, the highest accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

[6] The research work done by Pushpendu Ghosh, Ariel Neufeld, Jajati Keshari SahooDepartment of Computer Science & Information Systems, BITS Pilani K.K.Birla Goa campus, India bDivision of Mathematical Sciences, Nanyang Technological University, Singapore cDepartment of Mathematics, BITS Pilani K.K.Birla Goa campus, India. We employ both random forests and

LSTM networks (more precisely CuDNNLSTM) as training methodologies to analyse their effectiveness in forecasting outof-sample directional movements of constituent stocks of the S&P 500 from January 1993 till December 2018 for intraday trading. We introduce a multi-feature setting consisting not only of the returns with respect to the closing prices, but also with respect to the opening prices and intraday returns. As trading strategy, we use Krauss et al. (2017) and Fischer & Krauss (2018) as benchmark and, on each trading day, buy the 10 stocks with the highest probability and sell short the 10 stocks with the lowest probability to outperform the market in terms of intraday returns – all with equal monetary weight. Our empirical results show that the multi-feature setting provides a daily return, prior to transaction costs, of 0.64% using LSTM networks, and 0.54% using random forests. Hence, we outperform the singlefeature setting in Fischer & Krauss (2018) and Krauss et al. (2017) consisting only of the daily returns with respect to the closing prices, having corresponding daily returns of 0 .41% and of 0 .39% with respect to LSTM and random forests, respectively. 1 Keywords: Random forest, LSTM, Forecasting, Statistical Arbitrage, Machine learning, Intraday trading.

3 PROBLEM FORMULATION

Everyone want to be rich in his life with low efforts and great advantages. Similarly, we want to look in our future with inner most desire as we do not want to take risks or we want to decrease risk factor. Stock market is a place where selling and purchasing can provide future aims of life. Now the question is that how we can get advantages from stock market? Or what are the steps that can give us stocks market predictions before taking yourself in risk zone

The stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short term price of an individual stock.

Previous methods of stock predictions involve the use of Artificial Neural Networks and Convolution Neural Networks which has an error loss at an average of 20%.

In this report, we will see if there is a possibility of devising a model using Recurrent Neural Network which will predict stock price with a less percentage of error. And if the answer turns to be YES, we will also see how reliable and efficient will this model be.

Fig. 3.1: IBM stock graph

3.1 TECHNOLOGIES

3.1.1 PYTHON

Python was the language of choice for this project. This was an easy decision for the multiple reasons.

1. Python as a language has an enormous community behind it. Any problems that might be encountered can be easily solved with a trip to Stack Overflow. Python is among the most popular languages on the site which makes it very likely there will be a direct answer to any query.

2. Python has an abundance of powerful tools ready for scientific computing. Packages such as Numpy, Pandas, and SciPy are freely available and well documented. Packages such as these can dramatically reduce, and simplify the code needed to write a given program. This makes iteration quick.

3. Python as a language is forgiving and allows for programs that look like pseudo code. This is useful when pseudocode given in academic papers needs to be implemented and tested. Using Python, this step is usually reasonably trivial.

However, Python is not without its flaws. The language is dynamically typed and packages are notorious for Duck Typing. This can be frustrating when a package method returns something that, for example, looks like an array rather than being an actual array. Coupled with the fact that standard Python documentation does not explicitly state the return type of a method, this can lead to a lot of trials and error testing that would not otherwise happen in a strongly typed language. This is an issue that makes learning to use a new Python package or library more difficult than it otherwise could be.

3.1.2 NUMPY

Numpy is python modules which provide scientific and higher level mathematical abstractions wrapped in python. In most of the programming languages, we can’t use mathematical abstractions such as f(x) as it would affect the semantics and the syntax of the code. But by using Numpy we can exploit such functions in our code.

Numpy’s array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy also provides basic numerical routines, such as tools for finding Eigenvectors.

3.1.3 SCIKIT LEARN

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machine, random forest, gradient boosting, k-means etc. It is mainly designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-learn is largely written in Python, with some core algorithms written in Cython to achieve performance. Support vector machines are implemented by a Cython wrapper around LIBSVM .i.e., logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR.

3.1.4 TENSORFLOW

TensorFlow is an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

TensorFlow is Google Brain's second-generation system. While the reference implementation runs on single devices, TensorFlow can run on multiple CPUs and GPUs (with optional CUDA and SYCL extensions for general-purpose computing on graphics processing units). TensorFlow is available on 64-bit Linux, macOS, Windows, and mobile computing platforms including Android and iOS.

3.1.5 KERAS

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

Keras allows for easy and fast prototyping (through user friendliness, modularity, and extensibility). Supports both convolutional networks and recurrent networks, as well as combinations of the two. Runs seamlessly on CPU and GPU.

The library contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on GitHub, and community support forums include the GitHub issues page, a Gitter channel and a Slack channel.

3.1.6 COMPILER OPTION

Anaconda is a freemium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system conda.

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short term memory (LSTM).

In order to predict the stock prices in future markets, we have analyzed papers and has given an overview on how these algorithms give precise and accurate future predictions. In this paper, we used several algorithms from which we observed that not all the algorithms implemented can predict data we need. There has been a basic requirement for computerized and automized ways to deal with powerful and proficient usage of huge measure of money related information to help organizations and people in vital arranging and decision making on investments.

4 RESEARCH OBJECTIVES

The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades.

There are two types of stocks. You may know of intraday trading by the commonly used term "day trading." Interday traders hold securities positions from at least one day to the next and often for several days to weeks or months. LSTMs are very powerful in sequence prediction problems because they're able to store past information.

This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down.

The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the future price of a stock market index based on its historical price information. Artificial intelligence techniques have the ability to take into consideration financial system complexities and they are used as financial time series forecasting tools.

Our objective is to predict the future price and calculate the future growth of the company in the different time span. Then we analyze the prediction error for each company of different

sector. Based on that we conclude which time span is best for future prediction of that particular sector. We first predict the future closing price of 5 different companies from some pre-decided sectors with the help of LSTM. This prediction will be done on historical data & the future prediction will be done for 3 month, 6- month, 1 year & 3 years. In these four different time spans (3 & 6 months, 1 & 3 years), we calculate the growth of those companies. Then by analyzing the deviations of closing price for each time span, we took the resultant time span which has maximum growth, i.e. less error for the particular sector, e.g. companies A, B, C, D & E from a sector S1 has more growth in 3-months' time span of prediction then we draw an conclusion that for sector S1, our framework gives the best prediction for next 3-months for that particular sector. In our analysis, let's consider we are using the data for Months.

The main objective is to forecast the current market trends and could predict the stock pricesaccurately. We use LSTM recurrent neural networks to predict the stock prices accurately.

Stock market is one of the major fields that investors are dedicated to, thus stock market price trend prediction is always a hot topic for researchers from both financial and technical domains. In this research, our objective is to build a state-of-art prediction model for price trend prediction, which focuses on short-term price trend prediction.

Fig. 4.1: tesla stock graph

As concluded by Fama in [26], financial time series prediction is known to be a notoriously difficult task due to the generally accepted, semi-strong form of market efficiency and the high level of noise. Back in 2003, Wang et al. in [44] already applied artificial neural networks on stock market price prediction and focused on volume, as a specific feature of stock market. One of the key findings by them was that the volume was not found to be effective in improving the forecasting performance on the datasets they used, which was S&P 500 and DJI. Ince and Trafalis in [15] targeted short-term forecasting and applied support vector machine (SVM) model on the stock price prediction. Their main contribution is performing a comparison between multi-layer perceptron (MLP) and SVM then found that most of the scenarios SVM outperformed MLP, while the result was also affected by different trading strategies. In the meantime, researchers from financial domains were applying conventional statistical methods and signal processing techniques on analyzing stock market data.

The optimization techniques, such as principal component analysis (PCA) were also applied in short-term stock price prediction [22]. During the years, researchers are not only focused on stock price-related analysis but also tried to analyze stock market transactions such as volume burst risks, which expands the stock market analysis research domain broader and indicates this research domain still has high potential [39]. As the artificial intelligence techniques evolved in recent years, many proposed solutions attempted to combine machine learning and deep learning techniques based on previous approaches, and then proposed new metrics that serve as training features such as Liu and Wang [23]. This type of previous works belongs to the feature engineering domain and can be considered as the inspiration of feature extension ideas in our research. Liu et al. in [24] proposed a convolutional neural network (CNN) as well as a long short-term memory (LSTM) neural network based model to analyze different quantitative strategies in stock markets. The CNN serves for the stock selection strategy, automatically extracts features based on quantitative data, then follows an LSTM to preserve the time-series features for improving profits.

The latest work also proposes a similar hybrid neural network architecture, integrating a convolutional neural network with a bidirectional long short-term memory to predict the stock market index [4]. While the researchers frequently proposed different neural network solution architectures, it brought further discussions about the topic if the high cost of training such models is worth the result or not.

There are three key contributions of our work (1) a new dataset extracted and cleansed (2) a comprehensive feature engineering, and (3) a customized long short-term memory (LSTM) based deep learning model.

We have built the dataset by ourselves from the data source as an open-sourced data API called Tushare [43]. The novelty of our proposed solution is that we proposed a feature engineering along with a fine-tuned system instead of just an LSTM model only. We observe from the previous works and find the gaps and proposed a solution architecture with a comprehensive feature engineering procedure before training the prediction model. With the success of feature extension method collaborating with recursive feature elimination algorithms, it opens doors for many other machine learning algorithms to achieve high accuracy scores for short-term price trend prediction. It proved the effectiveness of our proposed feature extension as feature engineering. We further introduced our customized LSTM model and further improved the prediction scores in all the evaluation metrics. The proposed solution outperformed the machine learning and deep learning-based models in similar previous works.

As one of the most popular financial management methods, stocks have attracted more and more investors to participate. The risks of stock investment are relatively high. How to reduce risks and increase profits has become the most concerned issue for investors. Traditional stock forecasting models use forecasting models based on stock time series analysis, but time series models cannot consider the influence of investor sentiment on stock market changes. In order to use investor sentiment information to make more accurate stock market forecasts, this paper establishes a stock index forecast and network security model based on time series and deep learning.

Based on the time series model, it is proposed to use CNN to extract in-depth emotional information to replace the basic emotional features of the emotional extraction level. At the data source level, other information sources, such as basic features, are introduced to further improve the predictive performance of the model. The results show that the algorithm is feasible and effective and can better predict the changes in the market stock index. This also proves that multiple information sources can improve the accuracy of model prediction more effectively than a single information source.

Stock price prediction has great value in seeking to maximize the profit of a stock investment, and related technologies have been studied for decades. According to the efficient market hypothesis, news can have an impact on stock prices, which also shows that events have a driving effect on the stock market. In the field of natural language processing (NLP), public news and social media are the two main data sources for stock prediction

The research work done by Mariam Moukalled Wassim El-Hajj Mohamad Jaber Computer Science Department American University of Beirut.

Traditionally and in order to predict market movement, investors used to analyse the stock

prices and stock indicators in addition to the news related to these stocks. Hence, the importance of news on the stock price movement. Most of the previous work in this industry

focused on either classifying the released market news as (positive, negative, neutral) and demonstrating their effect on the stock price or focused on the historical price movement and predicted their future movement. In this work, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news' sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted, the highest accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

5 METHODOLOGY

PROPOSED SYSTEMS

The prediction methods can be roughly divided into two categories, statistical methods and artificial intelligence methods. Statistical methods include logistic regression model, ARCH model, etc. Artificial intelligence methods include multi-layer perceptron, convolutional neural network, naive Bayes network, back propagation network, single-layer LSTM, support vector machine, recurrent neural network, etc. They used Long short-term memory network (LSTM).

Long short-term memory network:

Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

Working of LSTM:

LSTM is a special network structure with three “gate” structures. Three gates are placed in an LSTM unit, called input gate, forgetting gate and output gate. While information enters the LSTM’s network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate.

The experimental data in this paper are the actual historical data downloaded from the Internet. Three data sets were used in the experiments. It is needed to find an optimization algorithm that requires less resources and has faster convergence speed.

- Used Long Short-term Memory (LSTM) with embedded layer and the LSTM neural network with automatic encoder.
- LSTM is used instead of RNN to avoid exploding and vanishing gradients.
- In this project python is used to train the model, MATLAB is used to reduce dimensions of the input. MySQL is used as a dataset to store and retrieve data.
- The historical stock data table contains the information of opening price, the highest price, lowest price, closing price, transaction date, volume and so on.
- The accuracy of this LSTM model used in this project is 57%.

LMS filter:

The LMS filter is a kind of adaptive filter that is used for solving linear problems. The idea of the filter is to minimize a system (finding the filter coefficients) by minimizing the least mean square of the error signal.

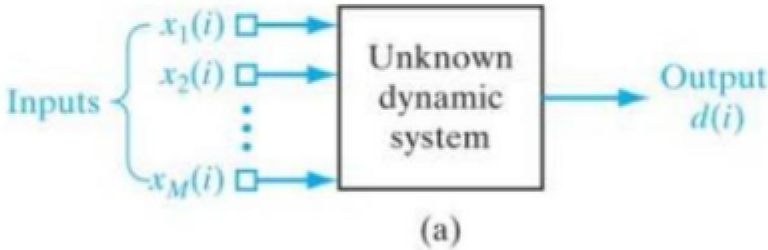

fig: 5.1 LMS input and output

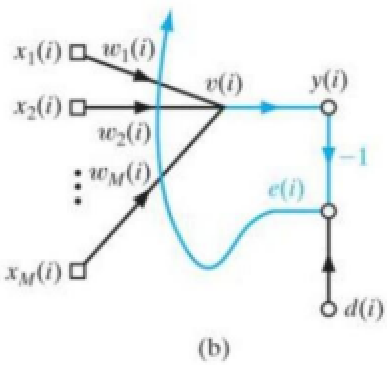

fig 5.2: LMS updating weights

Algorithm 1: LMS

Input:
x : input vector
d: desired vector
μ: learning rate
N: filter order
Output:
y: filter response
e: filter error
begin
  $M = size(x)$;
  $x_n(0) = w_n(0) = [0\ 0\ \dots\ 0]^T$;
  while $n < M$ **do**
    $x_{n+1} = [x(n);\ x_n(1:N)]$;
    $y(n) = w_n^H * x_n$;
    $e(n) = d(n) - y(n)$;
    $w_{n+1} = w_n + 2\mu e(n) x_n$;
  end
end

In general, we don't know exactly if the problem can be solved very well with linear approach, so we usually test a linear and a non-linear algorithm. Since the internet always shows non-linear approaches, we will use LMS to prove that stock market prediction can be done with linear algorithms with a good precision.

But this filter mimetizes a system, that is, if we apply this filter in our data, we will have the filter coefficients trained, and when we input a new vector, our filter coefficients will output a response that the original system would (in the best case). So we just have to do a tricky modification for using this filter to predict data.

The system:

First, we will delay our input vector by l positions, where l would be the quantity of days we want to predict, this l new positions will be filled by zeros.

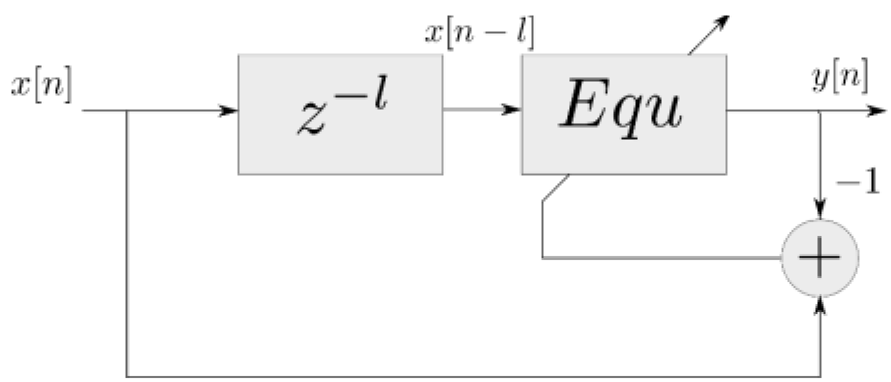

Fig. 5.3: LMS updating weights

When we apply the LMS filter, we will train the filter to the first 178 data. After that, we will set the error as zero, so the system will start to output the answers as the original system to the last l values. We will call the tricky modification as the LMSPred algorithm.

Algorithm 2: LMSPred

Input:
x: input vector
l: quantity of days to predict
μ: learning rate
N: filter order
Output:
y: filter response

```
begin
    M = size(x_d);
    x_n(0) = w_n(0) = [0 0 ... 0];
    x_d = [0 0 .. 0 x];
    while n < M do
        x_{n+1} = [x_d(n); x_n(1 : N)];
        y(n) = w_n^H * x_n;
        if n > M - l then
            e = 0;
        else
            e(n) = d(n) - y(n);
        end
        w_{n+1} = w_n + 2μ e(n) x_n;
    end
end
```

LSTM Architecture

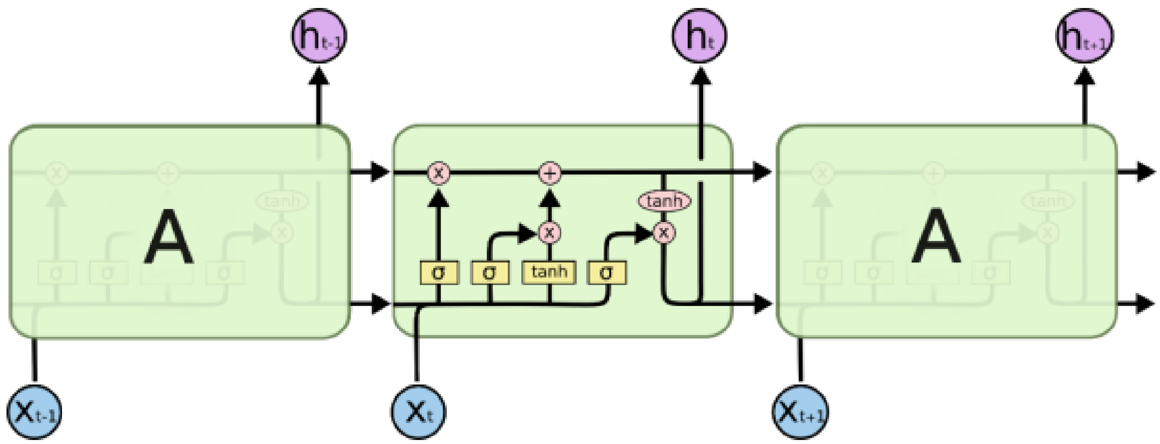

Fig 5.4: LSTM architecture

Forget Gate:

A forget gate is responsible for removing information from the cell state.

- The information that is no longer required for the LSTM to understand things or the information that is of less importance is removed via multiplication of a filter.
- This is required for optimizing the performance of the LSTM network.
- This gate takes in two inputs; h_t-1 and x_t. h_t-1 is the hidden state from the previous cell or the output of the previous cell and x_t is the input at that particular time step.

Input Gate:

1. Regulating what values need to be added to the cell state by involving a sigmoid function. This is basically very similar to the forget gate and acts as a filter for all the information from hi-1 and x_t.
2. Creating a vector containing all possible values that can be added (as perceived from h_t-1 and x_t) to the cell state. This is done using the tanh function, which outputs values from -1 to +1.
3. Multiplying the value of the regulatory filter (the sigmoid gate) to the created vector (the tanh function) and then adding this useful information to the cell state via addition operation.

Output Gate:

The functioning of an output gate can again be broken down to three steps:

- Creating a vector after applying tanh function to the cell state, thereby scaling the values to the range -1 to +1.

- Making a filter using the values of h_t-1 and x_t, such that it can regulate the values that need to be output from the vector created above. This filter again employs a sigmoid function.
- Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as a output and also to the hidden state of the next cell.

- # LSTM
- Inputs: dataset
- Outputs: RMSE of the forecasted data
-
- # Split dataset into 75% training and 25% testing data
- size = length(dataset) * 0.75
- train = dataset [0 to size]
- test = dataset [size to length(dataset)]
-
- # Procedure to fit the LSTM model
- Procedure LSTMAlgorithm (train, test, train_size, epochs)

- X = train
- y = test
- model = Sequential ()
- model.add(LSTM(50), stateful=True)
- model.compile(optimizer='adam', loss='mse')
- model.fit(X, y, epochs=epochs, validation_split=0.2)
- return model
-
- # Procedure to make predictions
- Procedure getPredictonsFromModel (model, X)
- predictions = model.predict(X) • return predictions
-
- epochs = 100
- neurons = 50
- predictions = empty 16
- # Fit the LSTM model
- model = LSTMAlgorithm (train, epoch, neurons)
-

- # Make predictions
- pred = model.predict(train)
-
- # Validate the model
- n = len(dataset)
-
- error = 0
- for i in range(n): error += (abs(real[i] - pred[i])/real[i]) * 100
- accuracy = 100 - error/n

Hardware Requirements:

- RAM: 4 GB
- Storage: 500 GB
- CPU: 2 GHz or faster
- Architecture: 32-bit or 64-bit

Software Requirements:

- Python 3.5 in Google Colab is used for data pre-processing, model training and prediction.

• Operating System: windows 7 and above or Linux based OS or MAC OS

Functional requirements

Functional requirements describe what the software should do (the functions). Think about the core operations.

Because the "functions" are established before development, functional requirements should be written in the future tense. In developing the software for Stock Price Prediction, some of the functional requirements could include:

• The software shall accept the tw_spydata_raw.csv dataset as input.

• The software should shall do pre-processing (like verifying for missing data values) on input for model training.

• The software shall use LSTM ARCHITECTURE as main component of the software.

• It processes the given input data by producing the most possible outcomes of a CLOSING STOCK PRICE.

Notice that each requirement is directly related to what we expect the software to do. They represent some of the core functions.

Non-Functional requirements Product properties

• Usability: It defines the user interface of the software in terms of simplicity of understanding the user interface of stock prediction software, for any kind of stock trader and other stakeholders in stock market.

• Efficiency: maintaining the possible highest accuracy in the closing stock prices in shortest time with available data.

Performance: It is a quality attribute of the stock prediction software that describes the responsiveness to various user interactions with it.

SYSTEM ARCHITECTURE

1) Preprocessing of data

Fig. 5.5: Pre-processing of data

2) Overall Architecture

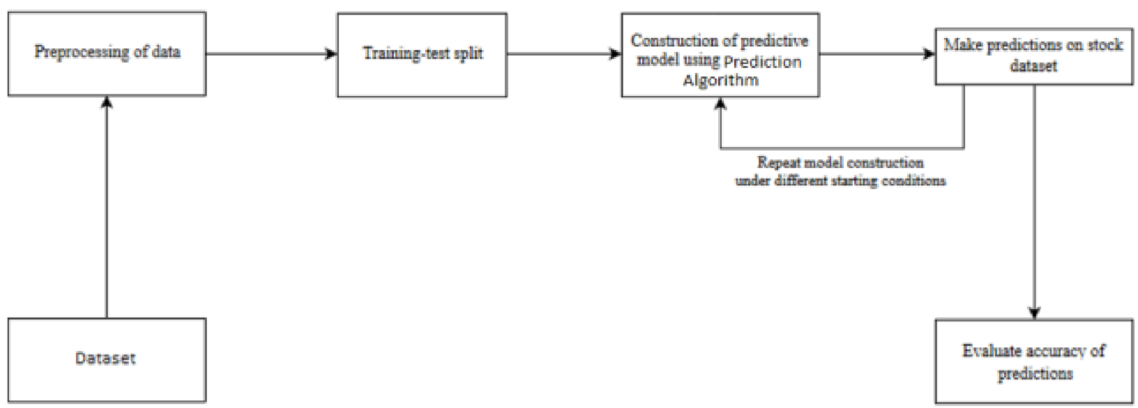

Fig. 5.6: Overall Architecture

This section explains the overall process of the literature collection on SMP usingmachine learning. Initially, the phrase “stock market prediction using machine learning”was keyed to various search engines, digital libraries and databases, including ‘google scholar’, ‘research gate’, ‘ACM digital library’, ‘IEEE Explore’, ‘Scopus’, and so on. Duringthe process of literature collection, various phrases like “stock market prediction methods”,“impact of sentiments on stock market prediction”, and “machine learning-based approachfor stock market prediction” were keyed. The OR and AND operators were used for thekeyword searches in single and multiple classes, respectively. As a result, some of thefundamental papers in the field of stock market prediction were retrieved. By the carefulanalysis of a few basic papers, a primary insight into the domain was obtained. The searchcriteria were further modified to collect the literature of the last decade, in order to enhanceand improve the domain. In addition, the literature selected was screened by applyingquality criteria, where metrics such as indexing, quartiles, impact factors and publisherswere observed.

We are using R statistical programming language to study and perform this experiment. -R is an extremely flexible statistics programming language that is Open Source and unreservedly accessible for all standard working frameworks. R has as of late encountered a "“explosive growth in use and in user contributed software". The "user-contributed software" is one of the most exceptional and gainful parts of R, as a huge number of clients have contributed code for actualizing probably the most avant-garde measurable strategies, notwithstanding R executing basically all standard measurable investigations. As a result of R's Open Source structure and a group of clients devoted to making R of the most noteworthy quality, the PC code on which the techniques are based is straight forwardly scrutinized and improved. By using R and implementing following Machine learning algorithms on the datasets we are predicting the stock price movement

6. RESULTS AND DISCUSSION

The proposed LSTM based model is implemented using Python.

WORKING:

1. Stock price prediction of Google

a) The Stock graph in real

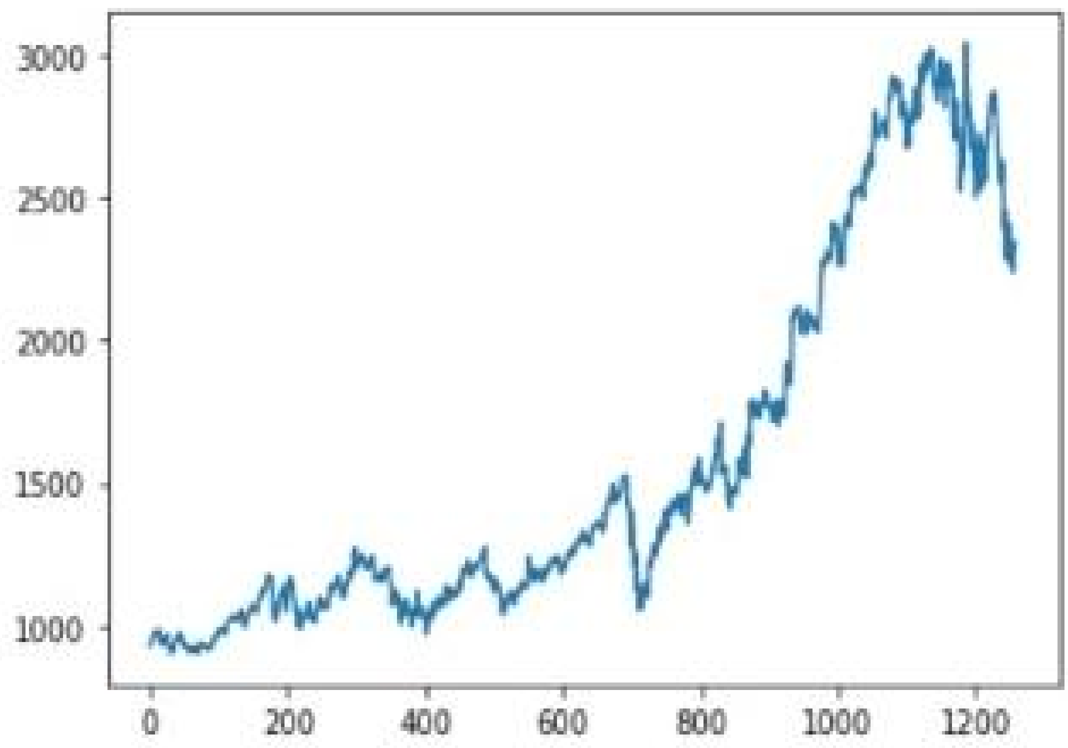

Fig 6.1 stock graph

b) Comparison of actual stock graph and predicted stock graph

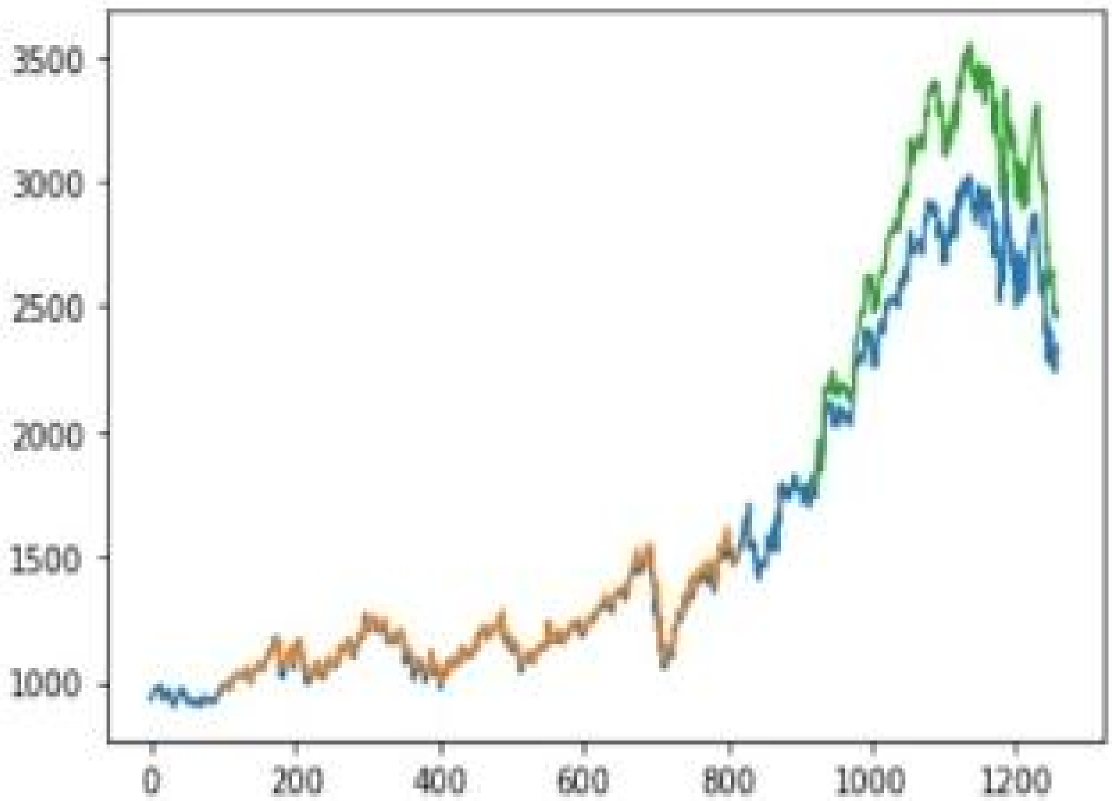

Fig. 6.2: stock graph comparison

c) Stock graph predicted

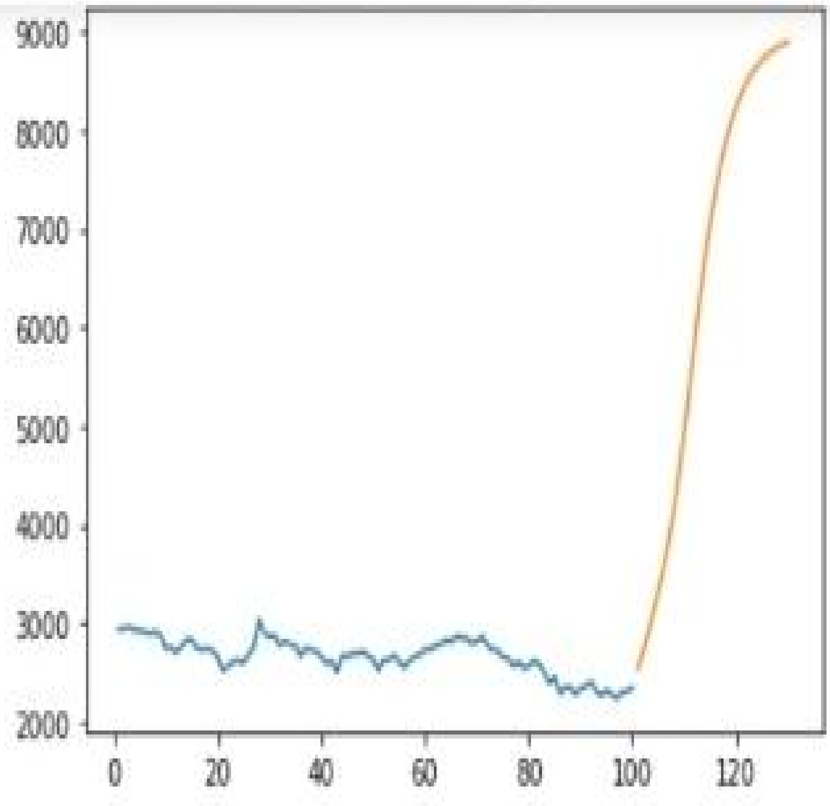

Fig. 6.3: Predicted graph

d) How future graph of stock will look

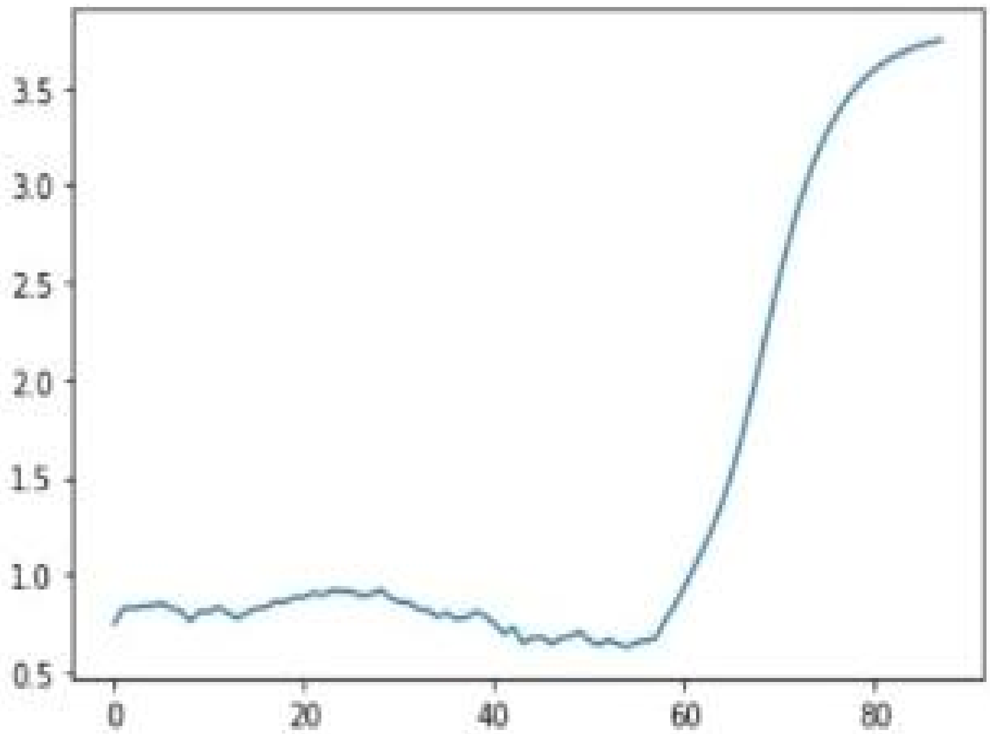

Fig. 6.4 Predicted graph

2. Stock price prediction of IBM

a) The Stock graph in real

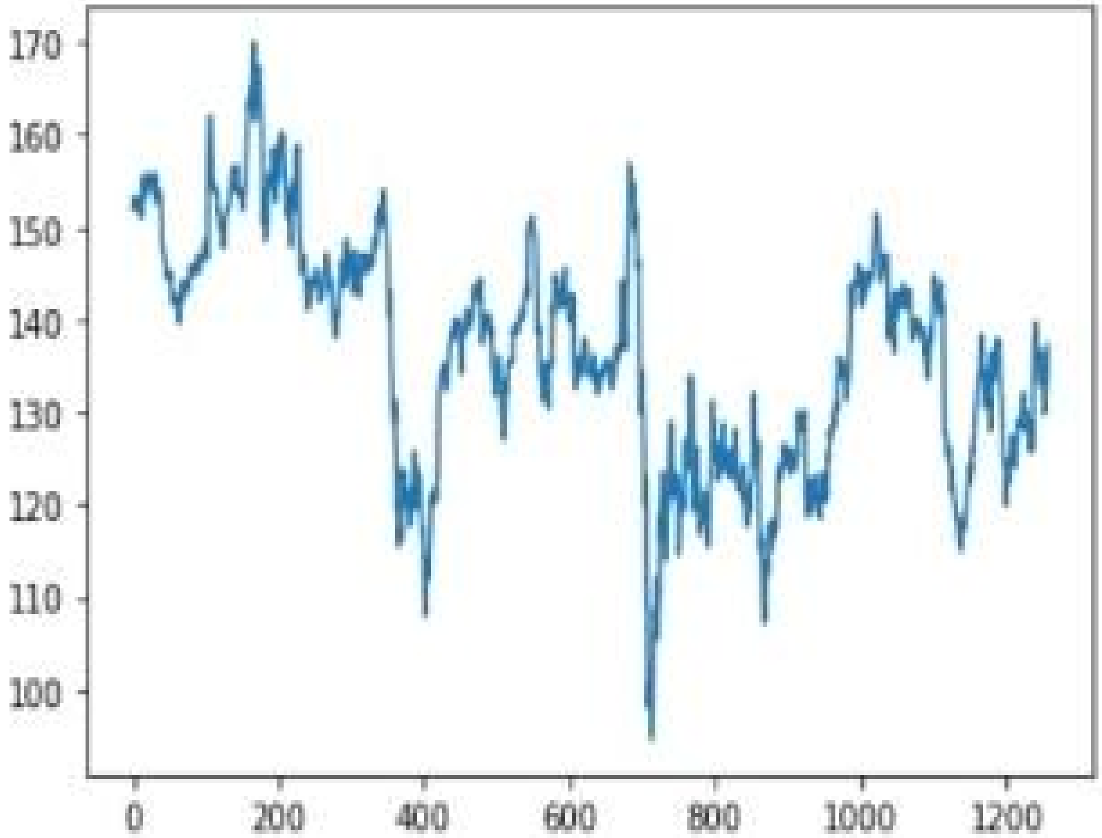

Fig 6.5 stock graph

b) Comparison of actual stock graph and predicted stock graph

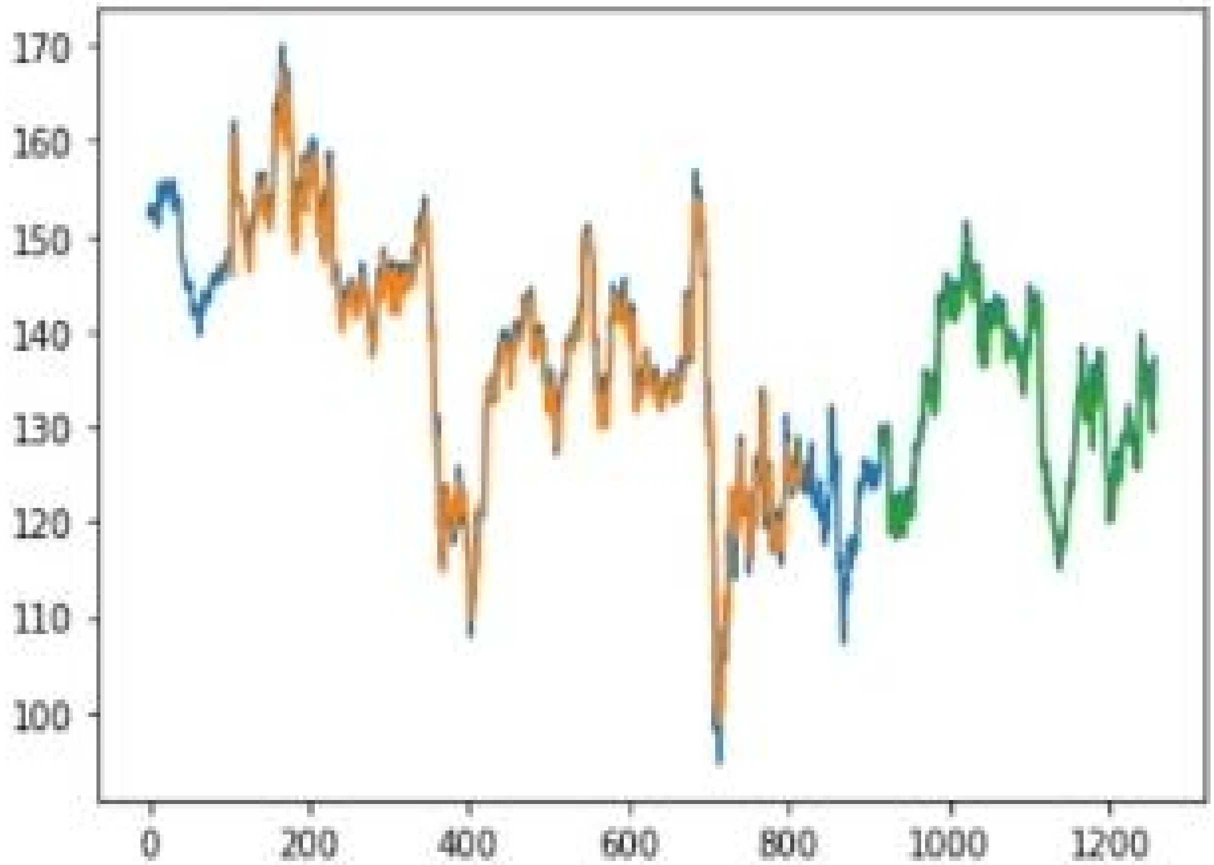

Fig. 6.6: stock graph comparison

c) Stock graph predicted

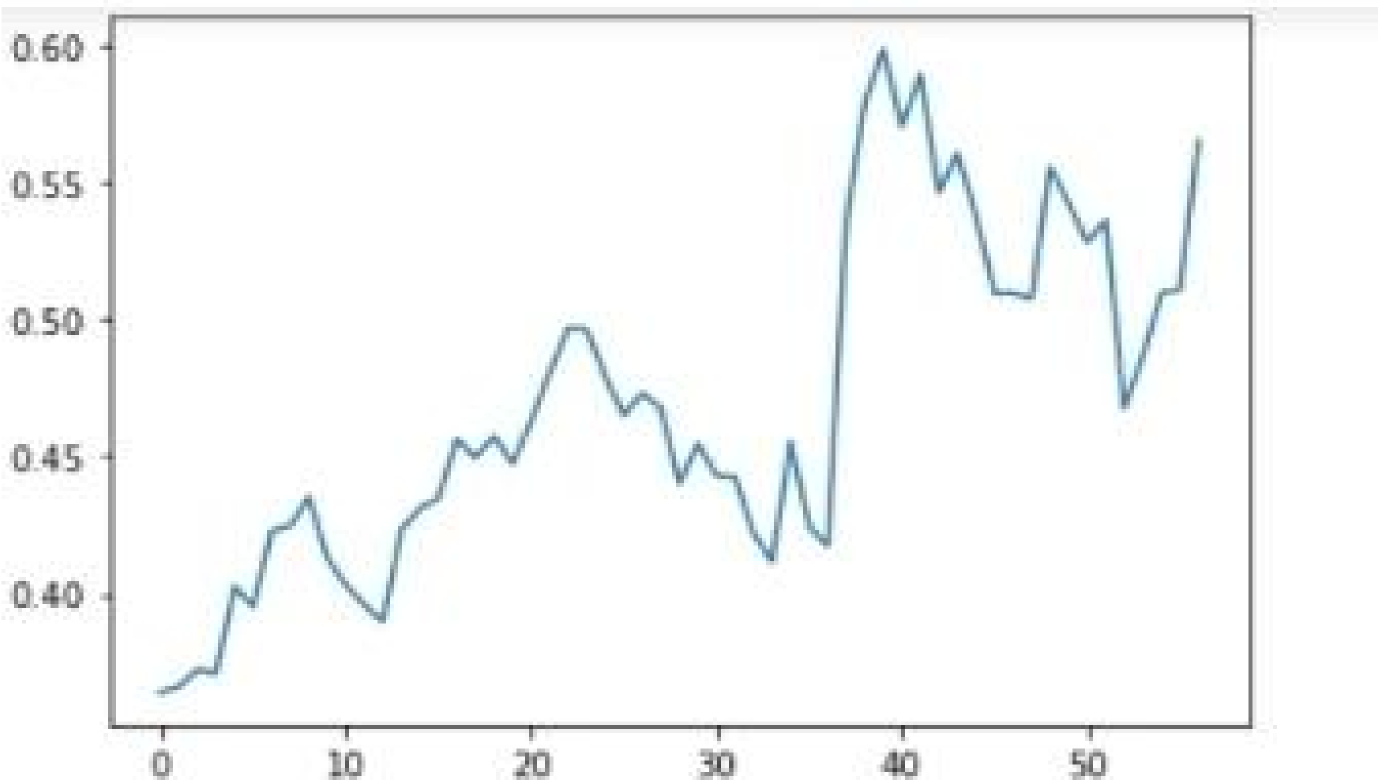

Fig. 6.7: Predicted graph

d) How future graph of stock will look

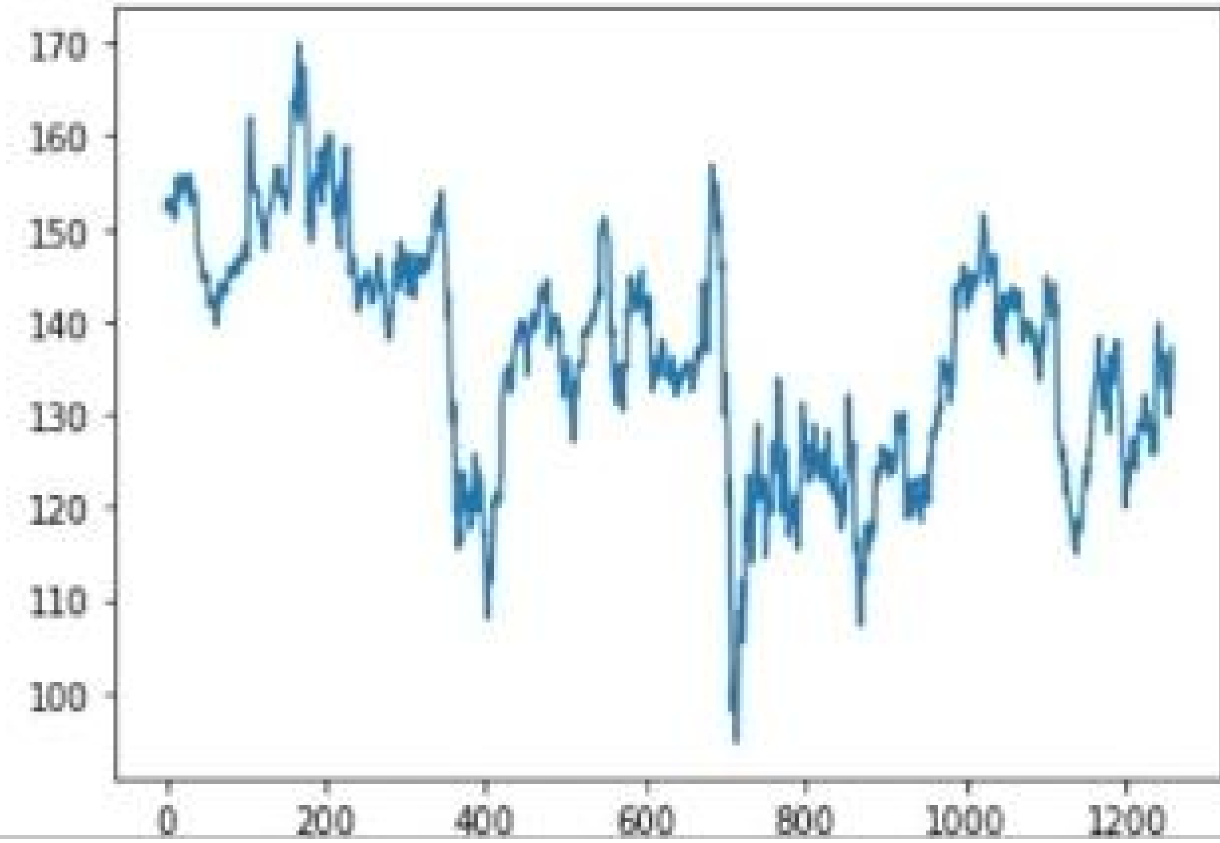

Fig. 6.8 Predicted graph

In this paper, we analyze the growth of the companies from different sector and try to find out which is the best time span for predicting the future price of the share. So, this draws an important conclusion that companies from a certain sector have the same dependencies as well as the same growth rate. The prediction can be more accurate if the model will train with a greater number of data set. Moreover, in the case of prediction of various shares, there may be some scope of specific business analysis. We can study the different pattern of the share price of different sectors and can analyze a graph with more different time span to fine tune the accuracy. This framework broadly helps in market analysis and prediction of growth of different companies in different time spans. Incorporating other parameters (e.g. investor sentiment, election outcome, geopolitical stability) that are not directly correlated with the closing price may improve the prediction accuracy.

From the research done so far it could be concluded that the RNN and LSTM libraries are very effective in determining the stock price trends effectively relative to the actual market trend. At the same time what we could find out is that the python libraries that were used as a part of the training process were not very optimal. As far as the training speed is considered the functions that we use from the mathematics principle have a lot faster speed comparatively and they consist of more detailed designs and significant improvements when tested under various situations. However, the python library functions are considered to be more adaptable. From our work done so far we can easily tell that certain stock trends can be predicted easily on the basis of certain general rules and regulations of the stock. This the main reason behind the existence of the private placement institutes. Few things such as optimization of the neural network parameters as well as the training process however always has much room for improvement. All these points would be considered as further steps in the research.

7. REFERENCES

1. Krishna, V. ScienceDirect ScienceDirect NSE Stock Stock Market Market Prediction Prediction Using Using Deep-Learning Deep-Learning Models Models. Procedia Comput. Sci. 2018, 132, 1351–1362.

2. Market Capitalization of Listed Domestic Companies (Current US$) Data. Available online: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD (accessed on 19 May 2021).

3. Upadhyay, A.; Bandyopadhyay, G. Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression. J. Bus. Stud. Q. 2012, 3, 16–39.

4. Tan, T.Z.; Quek, C.; Ng, G.S. Biological Brain-Inspired Genetic Complementary Learning for Stock Market and Bank Failure Prediction. Comput. Intell. 2007, 23, 236–261. [CrossRef]

5. Ali Khan, J. Predicting Trend in Stock Market Exchange Using Machine Learning Classifiers. Sci. Int. 2016, 28, 1363–1367.

6. Gupta, R.; Garg, N.; Singh, S. Stock Market Prediction Accuracy Analysis Using Kappa Measure. In Proceedings of the 2013

International Conference on Communication Systems and Network Technologies, Gwalior, India, 6–8 April 2013; pp. 635–639.

7. Fama, E.F. Random walks in stock-market prices. Financ. Anal. J. 1995, 51, 75–80. [CrossRef]

8. Bujari, A.; Furini, M.; Laina, N. On using cashtags to predict companies stock trends. In Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 8–11 January 2017; pp. 25–28.

9. Inthachot, M.; Boonjing, V.; Intakosum, S. Artificial Neural Network and Genetic Algorithm Hybrid Intelligence for Predicting Thai Stock Price Index Trend. Comput. Intell. Neurosci. 2016, 2016, 3045254. [CrossRef]

10. Park, C.-H.; Irwin, S.H. What Do We Know about the Profitability of Technical Analysis? J. Econ. Surv. 2007, 21, 786–826. [CrossRef]

11. Venkatesh, C.K.; Tyagi, M. Fundamental Analysis as a Method of Share Valuation in Comparison with Technical Analysis. Bangladesh Res. Publ. J. 2011, 1, 167–174.

12. Nair, B.B.; Mohandas, V. An intelligent recommender system for

stock trading. Intell. Decis. Technol. 2015, 9, 243–269. [CrossRef]

13. Shiller, R. Measuring bubble expectations and investor confidence RJ Shiller. J. Psychol. Financ. Mark. 2000, 1, 49–60. [CrossRef]

14. Aharon, D.Y.; Gavious, I.; Yosef, R. Stock market bubble effects on mergers and acquisitions. Q. Rev. Econ. Financ. 2010, 50, 456–470. [CrossRef]

15. Molodovsky, N. A Theory of Price-Earnings Ratios. Financ. Anal. J. 1953, 9, 65–80. [CrossRef]

16. Kurach, R.; Sło ´nski, T. The PE Ratio and the Predicted Earnings Growth—The Case of Poland. Folia Oecon. Stetin. 2015, 15, 127–138. [CrossRef]

17. Dutta, A. Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. Intern. J. Bus. Inf. 2012, 7, 105–136.

18. Deboeck, G. Trading on the Edge: Neural, Genetic, and Fuzzy

Systems for Chaotic Financial Markets; Wiley: New York, NY, USA, 1994.

19. Zhu, Y.; Zhou, G. Technical analysis: An asset allocation perspective on the use of moving averages. J. Financ. Econ. 2009, 92, 519–544. [CrossRef]

20. Peachavanish, R. Stock selection and trading based on cluster analysis of trend and momentum indicators. Lect. Notes Eng. Comput. Sci. 2016, 1, 317–321.

21. Hulbert, M. Viewpoint: More Proof for the Dow Theory. Available online: https://www.nytimes.com/1998/09/06/business/viewpoint-more-proof-for-the-dow-theory.html (accessed on 17 October 2021).

22. Rahman, A.S.A.; Abdul-Rahman, S.; Mutalib, S. Mining Textual Terms for Stock Market Prediction Analysis Using Financial News. In International Conference on Soft Computing in Data Science; Springer: Singapore, 2017; pp. 293–305. [CrossRef]

23. Ballings, M.; Poel, D.V.D.; Hespeels, N.; Gryp, R. Evaluating multiple classifiers for stock price direction prediction. Expert Syst. Appl. 2015, 42, 7046–7056. [CrossRef]

24. Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297. [CrossRef]

25. Srivastava, D.K.; Bhambhu, L. Data classification using support vector machine. J. Theor. Appl. Inf. Technol. 2010, 12, 1–7.

26. Venugopal, K.R.; Srinivasa, K.G.; Patnaik, L.M. Fuzzy based neuro—Genetic algorithm for stock market prediction. Stud. Comput. Intell. 2009, 190, 139–166.

27. Number of Social Network Users Worldwide from 2017 to 2025. Available online: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (accessed on 30 May 2021).

28. Ding, X.; Zhang, Y.; Liu, T.; Duan, J. Using Structured Events to Predict Stock Price Movement: An Empirical Investigation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1415–1425.

29. Howells, K.; Ertugan, A. Applying fuzzy logic for sentiment analysis of social media network data in marketing. Procedia Comput.

Sci. 2017, 120, 664–670. [CrossRef]

30. Liu, B. Sentiment Analysis and Opinion Mining; Morgan & Claypool Publishers: San Rafael, CA, USA, 2012; p. 167.

31. Li, W. Improvement of Stochastic Competitive Learning for Social Network. Comput. Mater. Contin. 2020, 63, 755–768.

32. Devi, K.N.; Bhaskaran, V.M. Semantic Enhanced Social Media Sentiments for Stock Market Prediction. Int. J. Econ. Manag. Eng. 2015, 9, 684–688.

33. Hill, S.; Ready-Campbell, N. Expert Stock Picker: The Wisdom of (Experts in) Crowds. Int. J. Electron. Commer. 2011, 15, 73–102. [CrossRef]

34. Chen, T.-L.; Chen, F.-Y. An intelligent pattern recognition model for supporting investment decisions in stock market. Inf. Sci. 2016, 346–347, 261–274. [CrossRef]

35. Ranco, G.; Aleksovski, D.; Caldarelli, G.; Grˇcar, M.; Mozetic, I. The effects of Twitter sentiment on stock price returns. PLoS ONE 2015, 10, e0138441. [CrossRef] [PubMed]

36. Bhardwaj, A.; Narayan, Y.; Dutta, M. Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty. Procedia Comput. Sci. 2015, 70, 85–91. [CrossRef]

37. Zhang, Y.; Wu, L. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. Expert Syst. Appl. 2009, 36, 8849–8854.

38. Guresen, E.; Kayakutlu, G.; Daim, T.U. Using artificial neural network models in stock market index prediction. Expert Syst. Appl. 2011, 38, 10389–10397. [CrossRef]

39. Lugmayr, A.; Gossen, G. Evaluation of methods and techniques for language based sentiment analysis for dax 30 stock exchange—A first concept of a 'LUGO' sentiment indicator. In Proceedings of the 5th International Workshop on Semantic Ambient Media Experience (SAME), Newcastle, UK, 18 June 2012; pp. 69–76.

40. Porshnev, A.; Redkin, I.; Karpov, N. Modelling Movement of Stock Market Indexes with Data from Emoticons of Twitter Users. Commun. Comput. Inf. Sci. 2015, 205, 297–306. [CrossRef]

41. Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A comprehensive evaluation of ensemble learning for stock-market prediction. J. Big Data 2020, 7, 1–40. [CrossRef]

42. Weng, B.; Ahmed, M.A.; Megahed, F. Stock market one-day ahead movement prediction using disparate data sources. Expert Syst. Appl. 2017, 79, 153–163. [CrossRef]

43. Di Persio, L.; Honchar, O. Recurrent neural networks approach to the financial forecast of Google assets. Int. J. Math. Comput. Simul. 2017, 11, 7–13.

44. Hagenau, M.; Liebmann, M.; Hedwig, M.; Neumann, D. Automated News Reading: Stock Price Prediction Based on Financial News Using Context-Specific Features. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–9 January 2012; pp. 1040–1049.

45. Huang, C.-F.; Li, H.-C. An Evolutionary Method for Financial Forecasting in Microscopic High-Speed Trading Environment. Comput. Intell. Neurosci. 2017, 2017, 9580815. [CrossRef] [PubMed]

46. Shah, D.; Isah, H.; Zulkernine, F. Stock Market Analysis: A Review

and Taxonomy of Prediction Techniques. Int. J. Financ. Stud. 2019, 7, 26. [CrossRef]

47. Nayak, A.; Pai, M.M.M.; Pai, R.M. Prediction Models for Indian Stock Market. Procedia Comput. Sci. 2016, 89, 441–449. [CrossRef]

48. Makrehchi, M.; Shah, S.; Liao, W. Stock Prediction Using Event-Based Sentiment Analysis. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 17–20 November 2013; Volume 1, pp. 337–342.

49. Ghanavati, M.; Wong, R.K.; Chen, F.; Wang, Y.; Fong, S. A Generic Service Framework for Stock Market Prediction. In Proceedings of the 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA, 27 June–2 July 2016; pp. 283–290.

50. Bustos, O.; Pomares, A.; Gonzalez, E. A comparison between SVM and multilayer perceptron in predicting an emerging financial market: Colombian stock market. In Proceedings of the 2017 Congreso

Internacional de Innovacion y Tendencias en Ingenieria
(CONIITI), Bogota, Colombia, 4–6 October 2017; pp. 1–6.

51. Dey, S.; Kumar, Y.; Saha, S.; Basak, S. Forecasting to Classification: Predicting the direction of stock market price using Xtreme
Gradient Boosting Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting.
PESIT South Campus 2016. [CrossRef]

52. Murshed, B.A.H.; Al-Ariki, H.D.E.; Mallappa, S. Semantic Analysis Techniques using Twitter Datasets on Big Data: Comparative
Analysis Study. Comput. Syst. Sci. Eng. 2020, 35, 495–512. [CrossRef]

53. Xie, B.; Passonneau, R.; Wu, L.; Creamer, G.G. Semantic Frames to Predict Stock Price Movement. In Proceedings of the 51st
Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; pp. 873–883.

54. Ding, X.; Zhang, Y.; Liu, T.; Duan, J. Knowledge-Driven Event Embedding for Stock Prediction. In Proceedings of the COLING

2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–17 December 2016; pp. 2133–2142.

55. Sirimevan, N.; Mamalgaha, I.G.U.H.; Jayasekara, C.; Mayuran, Y.S.; Jayawardena, C. Stock Market Prediction Using Machine Learning Techniques. In Proceedings of the IEEE 2019 International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 5–7 December 2019; Volume 1, pp. 192–197.

56. Schumaker, R.P.; Zhang, Y.; Huang, C.-N.; Chen, H. Evaluating sentiment in financial news articles. Decis. Support Syst. 2012, 53, 458–464. [CrossRef]

57. Huang, C.-J.; Liao, J.-J.; Yang, D.-X.; Chang, T.-Y.; Luo, Y.-C. Realization of a news dissemination agent based on weighted association rules and text mining techniques. Expert Syst. Appl. 2010, 37, 6409–6413. [CrossRef]

58. Nguyen, T.H.; Shirai, K.; Velcin, J. Sentiment analysis on social media for stock movement prediction. Expert Syst. Appl. 2015, 42, 9603–9611. [CrossRef]

59. Rajput, V.; Bobde, S. Stock market prediction using hybrid approach. In Proceedings of the 2016 International Conference on

Computing, Communication and Automation (ICCCA), Greater Noida, India, 29–30 April 2016; pp. 82–86.

60. Asghar, M.Z.; Subhan, F.; Imran, M.; Kundi, F.M.; Khan, A.; Shamshirband, S.; Mosavi, A.; Koczy, A.R.V.; Csiba, P. Performance Evaluation of Supervised Machine Learning Techniques for Efficient Detection of Emotions from Online Content. Comput. Mater. Contin. 2020, 63, 1093–1118. [CrossRef]

61. Akhtar, M.J.; Ahmad, Z.; Amin, R.; Almotiri, S.H.; Al Ghamdi, M.A.; Aldabbas, H. An Efficient Mechanism for Product Data Extraction from E-Commerce Websites. Comput. Mater. Contin. 2020, 65, 2639–2663. [CrossRef]

62. Pandarachalil, R.; Sendhilkumar, S.; Mahalakshmi, G.S. Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach. Cogn. Comput. 2014, 7, 254–262. [CrossRef]

63. Pagolu, V.S.; Reddy, K.N.; Panda, G.; Majhi, B. Sentiment analysis of Twitter data for predicting stock market movements. In Proceedings of the 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, 3–5 October 2016; pp. 1345–1350.

64. Mittal, A.; Goel, A. Stock Prediction Using Twitter Sentiment Analysis; Stanford University: Stanford, CA, USA, 2009; Volume 1, pp. 337–342.

65. Zhang, X.; Fuehres, H.; Gloor, P.A. Gloor, Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”. Procedia-Soc. Behav. Sci. 2011, 26, 55–62. [CrossRef]

66. Uysal, A.; Gunal, S. The impact of preprocessing on text classification. Inf. Process. Manag. 2014, 50, 104–112. [CrossRef]

67. Wang, J.; Wang, X.; Yang, Y.; Zhang, H.; Fang, B. A Review of Data Cleaning Methods for Web Information System. Comput. Mater. Contin. 2020, 62, 1053–1075. [CrossRef]

68. Zhong, X.; Enke, D. Forecasting daily stock market return using dimensionality reduction. Expert Syst. Appl. 2017, 67, 126–139. [CrossRef]

69. Ihlayyel, H.A.; Sharef, N.M.; Nazri, M.Z.A.; Abu Bakar, A. An enhanced feature representation based on linear regression model for stock market prediction. Intell. Data Anal. 2018, 22, 45–76.

[CrossRef]

70. Zhou, P.-Y.; Chan, K.C.C.; Ou, C.X. Corporate Communication Network and Stock Price Movements: Insights from Data Mining. IEEE Trans. Comput. Soc. Syst. 2018, 5, 391–402. [CrossRef]

71. Chandar, S.K. Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction. J. Ambient. Intell. Humaniz. Comput. 2019, 1–9. [CrossRef]

72. Sedighi, M.; Jahangirnia, H.; Gharakhani, M.; Fard, S.F. A Novel Hybrid Model for Stock Price Forecasting Based on Metaheuristics and Support Vector Machine. Data 2019, 4, 75. [CrossRef]

73. Khan, W.; Malik, U.; Ghazanfar, M.A.; Azam, M.A.; Alyoubi, K.H.; Alfakeeh, A. Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. Soft Comput. 2019, 24, 11019–11043. [CrossRef]

74. Ullah, K.; Qasim, M. Google Stock Prices Prediction Using Deep Learning. In Proceedings of the 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 9 November 2020; pp. 108–113.

75. Nassirtoussi, A.K.; Aghabozorgi, S.; Wah, T.Y.; Ngo, D.C.L. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. Expert Syst. Appl. 2015, 42, 306–324. [CrossRef]

76. Wu, D.D.; Olson, D.L. Enterprise Risk Management in Finance; Palgrave Macmillan: London, UK, 2015.

77. Chen, M.-Y.; Liao, C.-H.; Hsieh, R.-P. Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach. Comput. Hum. Behav. 2019, 101, 402–408. [CrossRef]

78. Das, S.R.; Mishra, D.; Rout, M. Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method. Expert Syst. Appl. X 2019, 4, 100016. [CrossRef]

79. Bouktif, S.; Fiaz, A.; Awad, M. Augmented Textual Features-Based

Stock Market Prediction. IEEE Access 2020, 8, 40269–40282.
[CrossRef]

80. Chen, S.; Zhou, C. Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network.
IEEE Access 2020, 9, 9066–9072. [CrossRef]

81. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations ofwords and phrases and their composi tionality. In Advances in Neural Information Processing Systems; The MIT Press: Cambridge, MA, USA, 2013; pp. 3111–3119.

82. Si, J.; Mukherjee, A.; Liu, B.; Li, Q.; Li, H.; Deng, X. Exploiting Topic based Twitter Sentiment for Stock Prediction. In Proceedings
of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; pp. 24–29.

83. Zhang, X.; Zhang, Y.; Wang, S.; Yao, Y.; Fang, B.; Yu, P.S. Improving stock market prediction via heterogeneous information fusion.
Knowl.-Based Syst. 2018, 143, 236–247. [CrossRef]

84. Zhou, Z.; Qin, J.; Xiang, X.; Tan, Y.; Liu, Q.; Xiong, N.N. News Text Topic Clustering Optimized Method Based on TF-IDF
Algorithm on Spark. Comput. Mater. Contin. 2020, 62, 217–231.
[CrossRef]

85. El Seidy, E.; Ibrahim, B.; Jamous, R.A.; Bayoum, B.I. A Novel Efficient Forecasting of Stock Market Using Particle Swarm Optimization with Center of Mass Based Technique. Int. J. Adv. Comput. Sci. Appl. 2016, 7. [CrossRef]

86. He, S.; Li, Z.; Tang, Y.; Liao, Z.; Li, F.; Lim, S.-J. Parameters Compressing in Deep Learning. Comput. Mater. Contin. 2020, 62, 321–336. [CrossRef]

87. Pestov, V. Is the k-NN classifier in high dimensions affected by the curse of dimensionality? Comput. Math. Appl. 2013, 65, 1427–1437. [CrossRef]

88. Kalra, V.; Aggarwal, R. Importance of Text Data Preprocessing & Implementation in RapidMiner. Proc. First Int. Conf. Inf. Technol. Knowl. Manag. 2018, 14, 71–75. [CrossRef]

89. Zhang, C.; Cheng, J.; Tang, X.; Sheng, V.S.; Dong, Z.; Li, J. Novel DDoS Feature Representation Model Combining Deep Belief

Network and Canonical Correlation Analysis. Comput. Mater. Contin. 2019, 61, 657–675. [CrossRef]

90. Sharma, S.; Ahmed, S.; Naseem, M.; Alnumay, W.S.; Singh, S.; Cho, G.H. A Survey on Applications of Artificial Intelligence for Pre-Parametric Project Cost and Soil Shear-Strength Estimation in Construction and Geotechnical Engineering. Sensors 2021, 21, 463. [CrossRef] [PubMed]

91. Ganser, A.; Hollaus, B.; Stabinger, S. Classification of Tennis Shots with a Neural Network Approach. Sensors 2021, 21, 5703. [CrossRef]

92. Ticknor, J.L. A Bayesian regularized artificial neural network for stock market forecasting. Expert Syst. Appl. 2013, 40, 5501–5506. [CrossRef]

93. Adebiyi, A.A.; Adewumi, A.; Ayo, C.K. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. J. Appl. Math. 2014, 2014, 614342. [CrossRef]

94. Chopra, S.; Yadav, D.; Chopra, A.N. Artificial Neural Networks Based Indian Stock Market Price Prediction: Before and After Demonetization. Int. J. Swarm Intell. Evol. Comput. 2019, 8, 174.

95. Seo, M.; Kim, G. Hybrid Forecasting Models Based on the Neural Networks for the Volatility of Bitcoin. Appl. Sci. 2020, 10, 4768. [CrossRef]

96. Vanstone, B.; Finnie, G.; Hahn, T. Creating trading systems with fundamental variables and neural networks: The Aby case study. Math. Comput. Simul. 2012, 86, 78–91. [CrossRef]

97. Khashei, M.; Hajirahimi, Z. Performance evaluation of series and parallel strategies for financial time series forecasting. Financ. Innov. 2017, 3, 24. [CrossRef]

98. Goh, A. Back-propagation neural networks for modeling complex systems. Artif. Intell. Eng. 1995, 9, 143–151. [CrossRef]

99. Zhang, F.; Li, J.; Wang, Y.; Guo, L.; Wu, D.; Wu, H.; Zhao, H. Ensemble Learning Based on Policy Optimization Neural Networks for Capability Assessment. Sensors 2021, 21, 5802. [CrossRef] [PubMed]

100. Bing, Y.; Hao, J.K.; Zhang, S.C. Stock Market Prediction Using Artificial Neural Networks. Adv. Eng. Forum 2012, 6, 1055–1060. [CrossRef]

101. Hota, H.S.; Handa, R.; Shrivas, A.K. Time Series Data Prediction Using Sliding Window Based RBF Neural Network. Int. J. Comput. Intell. Res. 2017, 13, 1145–1156.

102. Guo, Z.; Ye, W.; Yang, J.; Zeng, Y. Financial index time series prediction based on bidirectional two dimensional locality preserving projection. In Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 10–12 March 2017; pp. 934–938.

103. Milosevic, N. Equity forecast: Predicting long term stock price movement using machine learning. arXiv 2016, arXiv:1603.00751.

104. Li, X.; Xie, H.; Wang, R.; Cai, Y.; Cao, J.; Wang, F.; Min, H.; Deng, X. Empirical analysis: Stock market prediction via extreme

learning machine. Neural Comput. Appl. 2014, 27, 67–78. [CrossRef]

105. More, A.M.; Rathod, P.U.; Patil, R.H.; Sarode, D.R.; Student, B. Stock Market Prediction System using Hadoop. Int. J. Eng. Sci. Comput. 2018, 8, 16138–16140.

106. Mohan, S.; Mullapudi, S.; Sammeta, S.; Vijayvergia, P.; Anastasiu, D.C. Stock Price Prediction Using News Sentiment Analysis. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 4–9 April 2019; pp. 205–208.

107. Xianya, J.; Mo, H.; Haifeng, L. Stock Classification Prediction Based on Spark. Procedia Comput. Sci. 2019, 162, 243–250. [CrossRef]

108. Yu, Y.; Duan, W.; Cao, Q. The impact of social and conventional media on firm equity value: A sentiment analysis approach. Decis. Support Syst. 2013, 55, 919–926. [CrossRef]

109. Lin, L.; Cao, L.; Wang, J.; Zhang, C. The applications of genetic algorithms in stock market data mining optimization. Manag. Inf. Syst. 2004, 10, 273–280.

110. Pimenta, A.; Nametala, C.; Guimarães, F.G.; Carrano, E.G. An Automated Investing Method for Stock Market Based on Multiobjective Genetic Programming. Comput. Econ. 2017, 52, 125–144. [CrossRef]

111. Strader, T.J.; Rozycki, J.J.; Root, T.H.; Huang, Y.H.J. Machine Learning Stock Market Prediction Studies: Review and Research Directions. J. Int. Technol. Inf. Manag. 2020, 28, 63–83.

112. Kim, Y.; Ahn, W.; Oh, K.J.; Enke, D. An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms. Appl. Soft Comput. 2017, 55, 127–140. [CrossRef]

113. Nair, B.B.; Dharini, N.M.; Mohandas, V. A Stock Market Trend Prediction System Using a Hybrid Decision Tree-Neuro-Fuzzy System. In Proceedings of the 2010 International Conference on Advances in Recent Technologies in Communication and Computing, Kottayam, India, 16–17 October 2010; pp. 381–385.

114. Chandar, S.K. Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach. Clust. Comput. 2017, 22, 13159–13166. [CrossRef]

115. Chang, P.-C.; Wu, J.-L.; Lin, J.-J. A Takagi–Sugeno fuzzy model combined with a support vector regression for stock trading forecasting. Appl. Soft Comput. 2016, 38, 831–842. [CrossRef]

116. Yolcu, O.C.; Lam, H.-K. A combined robust fuzzy time series method for prediction of time series. Neurocomputing 2017, 247, 87–101. [CrossRef]

117. Rajab, S.; Sharma, V. An interpretable neuro-fuzzy approach to stock price forecasting. Soft Comput. 2019, 23, 921–936. [CrossRef]

118. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef] [PubMed]

119. Wu, H.; Liu, Y.; Wang, J. Review of Text Classification Methods on Deep Learning. Comput. Mater. Contin. 2020, 63, 1309–1321. [CrossRef]

120. Hoseinzade, E.; Haratizadeh, S. CNNpred: CNN-based stock market prediction using a diverse set of variables. Expert Syst. Appl. 2019, 129, 273–285. [CrossRef]

121. Gao, P.; Zhang, R.; Yang, X. The Application of Stock Index Price Prediction with Neural Network. Math. Comput. Appl. 2020, 25, 53. [CrossRef]

122. Sezer, O.; Ozbayoglu, A. Financial Trading Model with Stock Bar Chart Image Time Series with Deep Convolutional Neural Networks. Intell. Autom. Soft Comput. 2018. [CrossRef]

123. Pang, X.; Zhou, Y.; Wang, P.; Lin, W.; Chang, V. An innovative neural network approach for stock market prediction. J. Supercomput. 2018, 76, 2098–2118. [CrossRef]

124. Shah, D.; Campbell, W.; Zulkernine, F.H. A Comparative Study of LSTM and DNN for Stock Market Forecasting. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 4148–4155.

125. Li, X.; Yang, L.; Xue, F.; Zhou, H. Time series prediction of stock price using deep belief networks with intrinsic plasticity. In Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 1237–1242.

126. Zhang, J.; Teng, Y.-F.; Chen, W. Support vector regression with modified firefly algorithm for stock price forecasting. Appl. Intell. 2018, 49, 1658–1674. [CrossRef]

127. Zheng, J.; Fu, X.; Zhang, G. Research on exchange rate forecasting based on deep belief network. Neural Comput. Appl. 2017, 31, 573–582. [CrossRef]

128. Hushani, P. Using Autoregressive Modelling and Machine Learning for Stock Market Prediction and Trading. In Third International Congress on Information and Communication Technology; Springer: Singapore, 2018; pp. 767–774. [CrossRef]

129. Nguyen, D.H.D.; Tran, L.P.; Nguyen, V. Predicting Stock Prices Using Dynamic LSTM Models. Int. Conf. Appl. Inform. 2019, 6, 199–212. [CrossRef]

130. Zhang, L.; Zhang, L.; Teng, W.; Chen, Y. Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning. Procedia Comput. Sci. 2013, 17, 695–703. [CrossRef]

131. Cakra, Y.E.; Trisedya, B.D. Stock price prediction using linear regression based on sentiment analysis. In Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 10–11 October 2015; pp. 147–154.

132. Bhuriya, D.; Kaushal, G.; Sharma, A.; Singh, U. Stock market predication using a linear regression. In Proceedings of the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 20–22 April 2017; pp. 510–513.

133. Gururaj, V.; Shriya, V.R.; Ashwini, K. Stock market prediction using linear regression and support vector machines. Int. J. Appl. Eng. Res. 2019, 14, 1931–1934.

134. Enke, D.; Grauer, M.; Mehdiyev, N. Stock market prediction with Multiple Regression, Fuzzy type-2 clustering and neural networks. Procedia Comput. Sci. 2011, 6, 201–206. [CrossRef]

135. Kamley, S.; Jaloree, S.; Thakur, R. Multiple Regression: A Data

Mining Approach for Predicting the Stock Market Trends Based on Open, Close and High Price of The Month. Int. J. Comput. Sci. Eng. Inf. Technol. Res. 2013, 3, 173–180.

136. Yuan, J.; Luo, Y. Test on the Validity of Futures Market's High Frequency Volume and Price on Forecast. In Proceedings of the 2014 International Conference on Management of e-Commerce and e-Government, Shanghai, China, 31 October–2 November 2014; pp. 28–32.

137. Imran, K. Prediction of stock performance by using logistic regression model: Evidence from Pakistan Stock Exchange (PSX). AJER 2018, 8, 247–258. [CrossRef]

138. Meesad, P.; Rasel, R.I. Predicting stock market price using support vector regression. In Proceedings of the 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 17–18 May 2013.

139. Siew, H.L.; Nordin, M.J. Regression techniques for the prediction of stock price trend. In Proceedings of the 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, Malaysia, 10–12 September 2012; pp. 99–103.

140. Ananthi, M.; Vijayakumar, K. Stock market analysis using candlestick regression and market trend prediction (CKRM). J. Ambient. Intell. Humaniz. Comput. 2020, 12, 4819–4826. [CrossRef]

141. Cheng, C.; Xu, W.; Wang, J. A Comparison of Ensemble Methods in Financial Market Prediction. In Proceedings of the 2012 Fifth International Joint Conference on Computational Sciences and Optimization, Harbin, China, 23–26 June 2012; pp. 755–759.

142. Bisoi, R.; Dash, P. A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. Appl. Soft Comput. 2014, 19, 41–56. [CrossRef]

143. Nair, B.B.; Mohandas, V.P.; Nayanar, N.; Teja, E.S.R.; Vigneshwari, S.; Teja, K.V.N.S. A Stock Trading Recommender System Based on Temporal Association Rule Mining. SAGE Open 2015, 5, 2158244015579941. [CrossRef]

144. Hu, H.; Tang, L.; Zhang, S.; Wang, H. Predicting the direction of stock markets using optimized neural networks with Google Trends. Neurocomputing 2018, 285, 188–195. [CrossRef]

145. Rather, A.M. A Hybrid Intelligent Method of Predicting Stock Returns. Adv. Artif. Neural Syst. 2014, 2014, 1–7. [CrossRef]

146. Kalaivaani, P.C.D.; Thangarajan, R. Enhancing the Classification Accuracy in Sentiment Analysis with Computational Intelligence Using Joint Sentiment Topic Detection with MEDLDA. Intell. Autom. Soft Comput. 2020, 26, 71–79. [CrossRef]

147. Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. Int. J. Data Min. Knowl. Manag. Process. 2015, 5, 1–11. [CrossRef]

148. Huang, J.; Ling, C. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng. 2005, 17, 299–310. [CrossRef]

149. de Oliveira, F.A.; Nobre, C.N.; Zárate, L.E. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. Expert Syst. Appl. 2013, 40, 7596–7606. [CrossRef]

150. Khan, W.; Ghazanfar, M.A.; Azam, M.A.; Karami, A.; Aloubi, K.H.; Alfakeeh, A.S. Stock market prediction using machine

learning classifiers and social media, news. J. Ambient. Intell. Humaniz. Comput. 2020, 1–24. [CrossRef]

151. Bergmeir, C.; Hyndman, R.; Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Comput. Stat. Data Anal. 2018, 120, 70–83. [CrossRef]

152. Ghiassi, M.; Saidane, H.; Zimbra, D. A dynamic artificial neural network model for forecasting time series events. Int. J. Forecast. 2005, 21, 341–362. [CrossRef]

153. Binkowski, M.; Marti, G.; Donnat, P. Autoregressive convolutional neural networks for asynchronous time series. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Volume 2, pp. 933–945.

154. Baek, Y.; Kim, H.Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. Expert Syst. Appl. 2018, 113, 457–480. [CrossRef]

155. Zheng, H.; Zhou, Z.; Chen, J. RLSTM: A New Framework of Stock Prediction by Using Random Noise for Overfitting Prevention. Comput. Intell. Neurosci. 2021, 2021, 8865816. [CrossRef]

156. Robles-Granda, P.D.; Belik, I.V. A Comparison of Machine Learning Classifiers Applied to Financial Datasets. In Proceedings of the World Congress on Engineering and Computer Science 2010, San Francisco, CA, USA, 20–22 October 2010.

157. Popper, N. Knight Capital Says Trading Glitch Cost It $440 Million—The New York Times. Available online: https://dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/ (accessed on 20 July 2020).

158. Phillips, M. Nasdaq: Here's Our Timeline of the Flash Crash. Wall Str. J. 2010. Available online: https://www.wsj.com/articles/BL-MB-21942 (accessed on 30 May 2021).

159. Gul, S.; Khan, M.T.; Saif, N.; Rehman, S.U.; Roohullah, S. Stock Market Reaction to Political Events (Evidence from Pakistan). J. Econ. Sustain. Dev. 2013, 4, 165–175.

160. Suriani, N.S.; Hussain, A.; Zulkifley, M.A. Sudden Event

Recognition: A Survey. Sensors 2013, 13, 9966–9998. [CrossRef] [PubMed]

161. Huang, S.Y.B.; Lee, C.-J.; Lee, S.-C. Toward a Unified Theory of Customer Continuance Model for Financial Technology Chatbots. Sensors 2021, 21, 5687. [CrossRef] [PubMed]

162. Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. The rise of social bots. Commun. ACM 2016, 59, 96–104. [CrossRef]

163. Kalyanaraman, V.; Kazi, S.; Tondulkar, R.; Oswal, S. Sentiment Analysis on News Articles for Stocks. In Proceedings of the 2014 8th Asia Modelling Symposium, Taipei, Taiwan, 23–25 October 2014; pp. 10–15.

164. Seethalakshmi, R. Analysis of Stock Market Predictor Variables using Linear Regression Analysis. Int. J. Pure Appl. Math. 2020, 119, 369–378.