International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Analyzing the Impact of Global Influencing Features with Weighted Attention Model for Stock Market Forecasting

[1]Kush Bhushanwar, [2]Dr. Kuntal Barua

Submitted: 27/05/2023 **Revised**: 18/07/2023 **Accepted**: 29/07/2023

Abstract: The present volatility of the stock markets makes forecasting stock trends extremely challenging owing to several socio economic and political factors other than market trends. While machine learning models can be used to perform regression analysis based on historical data trends, it becomes extremely challenging to incorporate the variabilities which are non-numeric in nature. Some of the factors which govern the rise and fall of stock prices are socio economic conditions, trade wars, current pandemic situation and global market slowdown, reliability of a company among others. Hence, one of the most effective ways to incorporate these trends is analyzing public trends pertaining to the same. While public sentiments may not always be coherent with prevailing market trends, yet they often portray the existential trends in the market and opinions of the public regarding potential purchases of stocks of a particular company in a given time period. This paper presents an approach which is an amalgamation of deep nets with attention, and opinion mining for forecasting stock trends. The attention vector employed as an additional input computed on the moving average allows for current trend analysis along with opinion mining from public datasets to encompass both numeric data trends and non-numeric data parameters pertaining to global influencing features. The regression and forecasting accuracy have been computed on a diverse set of datasets to validate the performance of the proposed approach.

Keywords:- *Stock Market Forecasting, Global Influencing Features, Weighted Attention Score, Regularization Factor, Forecasting Accuracy.*

1. Introduction

Financial assessment and investing depend critically on stock market trend analysis. While stock market trend analysis is a time series regression problem, what makes it extremely complicated is the dependence on several non-numeric parameters such as socio economic conditions, political conditions of a country, political stability, financial crisis and trade wars, global slowdown and public sentiments pertaining to a company etc [1]. This leads to variabilities in the stock trends often exhibiting non-coherent patterns with respect to historical data [2].

Economic conditions can have a significant impact on stock trends. For example, during a recession, stock prices tend to decline as investors become more risk-averse. Conversely, during periods of economic growth, stock prices tend to rise as investors become more optimistic about the future [3]. Political factors can also impact stock trends. For example, changes in government policies, such as tax rates or regulations, can affect the performance of specific sectors of the economy and, in turn, impact stock prices [4]. The performance of individual companies can also impact stock trends. While conducive news updates such a rising profits and new product launches may result in sudden upward spikes in stock prices, negative news updates such as decline in sales or product recalls may have the opposite effect on the stock prices [5].

Global events, such as natural disasters, geopolitical tensions, or pandemics, can also impact stock trends [6]. These events can create uncertainty in the markets and cause investors to become more risk-averse and hence it is essential to note that stock trends are inherently variable and can be influenced by a wide range of factors [7]. It's also worth mentioning that past performance does not guarantee future results, so investors should exercise caution when making investment decisions based on historical stock trends [8]. The variabilities in stock trends are influenced by a variety of factors, including economic conditions, political factors, company performance, market sentiment, and global events [9]. While several approaches have been developed to predict stock market trends, yet it is often challenging to incorporate the non-numeric global influencing factors as a feature for stock trend analysis. One of the approaches which has been proven to be effective in incorporating global influencing factors along with public sentiments is using opinion mining and sentiment analysis for garnering non-numeric data as an exogenous input in addition to historical numeric data trends [10].

Various trend analysis techniques try to estimate the movement of stock trends based on the variables of influence. The most ubiquitous analysis methods to estimate future stock trends happen to be fundamental

1 Research Scholar, Institute of Engineering, SAGE University, Indore (M.P.), India
2 Associate Professor, Institute of Advance Computing, SAGE University, Indore (M.P.), India
1 kush.bhushanwar@gmail.com
2 kuntal.barua@gmail.com

analysis methods and technical analysis methods. Fundamental analysis methods involve studying the financial and economic factors that affect a company's stock price. Additionally, fundamental analysts consider macroeconomic factors such as interest rates, inflation, and GDP growth rates. The goal of fundamental analysis is to identify companies that are undervalued or overvalued based on their financial and economic metrics [11]. Technical analysis, on the other hand, involves analyzing past market data, such as stock prices and trading volumes, to identify patterns that can be used to predict future stock price movements.

White technical methods are inherently data centric in mature, it is essential to incorporate fundamental methods to gain future insights regarding global influencing features impacting stock trends. It is typically challenging to include the apt influences of macroeconomic features in terms of numerical features as they are extremely pervasive in natures and often exhibit bias or alignment towards particular stocks, in addition to being non-numeric inherently. Some of the noteworthy contemporary approaches in the domain are cited next. Kim et al. proposed an amalgamation of effective transfer entropy (ETE) for prediction of stock trends of benchmark S&P datasets [12]. Li et al. proposed preparing a training vector incorporating sentiment data from local news analysis an exogenous inputs [13]. Sen et al. proposed a GARCH model for stock market forecasting incorporating influencing factors for the Indian stock market [14]. Lobeila et al. proposed the LSTM and ARIMA models for estimating stock markets based on NASDAQ listed companies [15]. Sharaf et al. proposed fusing sentiment data related to COVID-19 with stacked LSTM models for stock market forecasting [16]. Li et al. proposed a statistical model employing Pearson Correlation Coefficient (PCC) to evaluate correlation among governing variables for estimating futures stock prices [17]. Duan et al. compared the performance of baseline deep learning models such as LSTM and bi-LST for evaluating the movement of stock market trends [18]. Ren et al. presented a combination of sentiment analysis and SVR model for stock market prediction. The approach used the sentiment polarities as an additional feature along with historical numerical data [19]. Jers et al. proposed the use of LSTM for time limited sliding windows for stock market forecasting with an aim to track the recent trends in the time series data over shorts spans [20]. Luo et al. presented a parallel stochastic gradient descent methods for stock market forecasting with a regularization parameter. The regularization parameter was shown to avoid sudden truncation of training at local minima and reduce iterations to convergence [21]. It can be observed that contemporary approaches try to incorporate both time series forecasting models along with methods to include global views and market trends for a more robust model design. The subsequent section presents the details of the proposed approach in terms of feature engineering, data filtration and model training.

2. Feature Engineering

Selection and preparation of training features for any data driven model is the first pivotal step for data driven models. This requires a judicious selection of training features which can fulfill the following objectives of the present research work, stated as [22]:

1) Incorporating global influencing features in terms of numeric values through opinion mining.

1) Offsetting effects of noise and disturbances inherent to stock datasets.

3) Employing a training algorithm which can render high prediction accuracy over a diverse set of benchmark datasets thereby validating the effectiveness of the prediction model [23].

2.1 Incorporating Global Influencing Features.

One of the most effective techniques to incorporate global influencing features is to extract intelligible information through sentiment analysis and opinion mining. It has been observed that stock market trends seems to have a clear reliance on public sentiments. However, public sentiments are often vague, biased and random in nature. Quantifying public sentiments is also challenging [24]. In this approach, the opinions of public pertaining to the stock market is gathered from twitter are quantified as tokens 13. The feature vector in this case is represented as:

$$\boldsymbol{T} = \begin{bmatrix} -\boldsymbol{K}; \boldsymbol{Negative\ Sentiments} \\ \boldsymbol{0}; \boldsymbol{Neutral\ Sentiments} \\ +\boldsymbol{K}; \boldsymbol{Positive\ Sentments} \end{bmatrix} \quad \textbf{(1)}$$

The depiction of sentiment polarities are shown in figure 1.

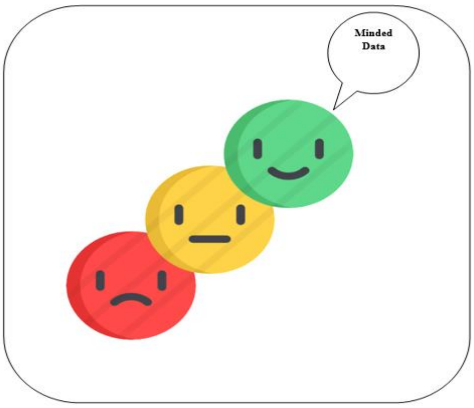

Fig.1 Sentiment Polarities

The tokens associated with sentiment polarities have been tokenized as, $-K, 0\ and +K$ respectively where the normalized value of K can be considered as, 1. Tokenization is often a fundamental step often incorporated in natural language processing (NLP) models aiming to extract opinions of sentiments from textual datasets. The goal of tokenization is to provide a structured representation of the text that can be easily processed by a machine learning model. Before analysing the sentiment of a piece of text, it needs to be divided into meaningful units [25]. Tokenization helps in standardizing the input text for further processing. Moreover, by breaking down text into tokens, the dimensionality of the input data is reduced. This can make it easier for machine learning algorithms to process the text efficiently [26]. Sentiment analysis usually involves classifying a piece of text (such as a review or a tweet) into one of several sentiment classes (e.g., positive, negative, neutral).

2.2 Filtering Baseline Noise

As the stock data sets are inherently noisy in nature, it is essential to filter out the baseline noise so as to aid pattern recognition over diverse datasets [27]. Stock data can be modelled as a time series function expressed as:

$$\boldsymbol{P_{stock} = f(Feature_{1-n}, Feature_{gi})} \qquad \textbf{(2)}$$

Here,

P_{stock} denotes the stock prices.

$Feature_{1-n}$ denotes the numeric features from $featrue\ 1\ to\ feature\ n$.

$Feature_{gi}$ denotes the global influencing features.

The time series representation of the stock prices can be transformed to the transform domain (typically frequency domain) to obtain multi resolution analysis of the data, which would in turn allow separating the low frequency information content and high frequency noise component [28]. The discrete wavelet transform (DWT) has been employed in this case for the same. The formulation for the DWT can be expressed as:

$$\boldsymbol{Z(x, c_{scale}, c_{shift}) = D^n \sum_i z(x) K_T \left[\frac{n - ic_{shift}}{c_{scale}}\right]} \ \textbf{(3)}$$

Here,

Z is the data in the transform domain.

$z(x)$ denotes the time series data.

c_{scale} denotes the scaling factor.

c_{shift} denotes the shifting factor.

n denotes the level of the DWT.

D^n denotes the dilation factor

K_T denotes the DWT family kernel.

The essence of the approach lies in the fact that the DWT separates the data into a high frequency con-efficient C_H often termed as the detailed co-efficient and a low frequency co-efficient C_L often termed as the approximate co-efficient. The high frequency component of the DWT decomposition is observed to contain the noise baseline and the low frequency component is observed to contain the maximum information content of the data [29] Thus retaining the C_L values over an iterative decomposition process, leaving out the C_H values results in effective noise filtering. While the flipside of the approach may be an evident decimation of the data samples, yet it could account for much more effective pattern recognition with respect to the original set of data samples [30].

2.3 Weighted Means

The next critically important aspect pertaining to feature preparation happens to create a feature vector which would identify the most recent trends in the data as an exogenous input apart from the overall data set [31]. This can be mathematically expressed as:

$$\boldsymbol{f_{mean}^{w} = \frac{1}{i-(k-l)}(f_i^v - \sum_k^l f_{l-k}^v)} \qquad \textbf{(4)}$$

Here,

f_{mean}^{w} denotes the weighted mean.

f_i^v denotes the instantaneous feature value.

k denotes the initial point for computing the mean.

l denotes the last point for computing the mean.

$\frac{1}{i-(k-l)}$ denotes the weight factor.

The essence of computing the weighted moving average as an separate feature is to emphasize the impact of most recent $(l-k)$ samples of the target variable. The term $\boldsymbol{\frac{1}{i-(k-l)}}$is used as a weight factor to keep the impact of the $(l-k)$ samples in bounds with respect to the total n chosen samples of the dataset [33]. Applying the weighted mean to both the input and target variables, the training feature set can be obtained as:

$$\boldsymbol{f_{composite} = \frac{1}{i-(k-l)}(f_i^v - \sum_k^l f_{l-k}^v)_{f1}, \frac{1}{i-(k-l)}(f_i^v - \sum_k^l f_{l-k}^v)_{f2} \dots \dots \frac{1}{i-(k-l)}(f_i^v - \sum_k^l f_{l-k}^v)_{fn},} \qquad \textbf{(5)}$$

Here,

$f_{composite}$ denotes the composite training vector.

Subscripts $1, 2 \dots . n$ denote the total number of independent feature variables.

The feature engineering serves as the prologue to the training algorithm for the designed deep neural net model, explained in the subsequent section [34].

3. Proposed Algorithm

The proposed algorithm essentially presents the deep neural network model with attention scores employing back propagation. The attention co-efficients are computed and used in conjugation with the back propagation model with an aim to achieve steepest descent [35].

If $\boldsymbol{s_i}$ denotes the states of the deep neural net, $\boldsymbol{r_{t-1}}$ denotes the regression constant, and $\boldsymbol{c_{a,i}}$, denotes the cosine alignment factor, then $\boldsymbol{c_{a,i}}$ is computed as:

$$\boldsymbol{c_{a,i} \xrightarrow[attention\ vector]{} r_{t-1}\ (s_i)} \quad \textbf{(6)}$$

The cosine alignment factor is computed to maximize the regression between the feature vector and the target variable. The cosine based overlap can be computed as:

$$\boldsymbol{C_{alignment} = \frac{1}{i-(k-l)} \int_i^n f_i^v(x) - \cos \sum_k^l f_{l-k}^v\,(x)_t . f_i^v(\tau - x - \sum_k^l f_{l-k}^v\,(\tau - x)_{\tau - x} dx} \quad \textbf{(7)}$$

The contextual attention vector can be further computed as:

$$\boldsymbol{A_{cont} = \sum_{i=k}^l C_{alignment} * r_{t-1}\ (s_i)_i} \quad \textbf{(8)}$$

The attention alignment vector as computed on the sliding sample window of $(l-k)samples$ enables the weighted average to be reinforced as an additional feature vector for the attention context thereby extracting the latest trends in time series stock dataset [36].

The proposed algorithm is presented as a sequential implementation of the steps put forth:

Proposed Algorithm:

Start

Step.1: Randomly divide data into training and testing sub-categories.

Step.2: Assign value of number of decomposition levels n ***and compute the DWT using*** $\boldsymbol{D^n \sum_i z(x) K_T \left[\frac{n - ic_{shift}}{c_{scale}}\right]}$.

Step.3: $\boldsymbol{for\ i = 1{:}n}$

{

retain values of $\boldsymbol{C_L}$ ***while discarding*** $\boldsymbol{C_H}$

}

Step.4: Assign values for $\boldsymbol{k\ and\ l}$ ***to compute the weighted mean*** $\boldsymbol{\frac{1}{i-(k-l)}(f_i^v - \sum_k^l f_{l-k}^v)}$.

Step.5: Compute the attention weights using the relation $\boldsymbol{\sum_{i=k}^l C_{alignment} * r_{t-1}\ (s_i)_i}$.

Step.6: Add the exogenous sentiment polarity vector $\boldsymbol{T \in [-K, 0, +K]}$.

Step.7: Define maximum number of iterations as $\boldsymbol{maxitr}$ ***and cost function to be minimized as:***

$$\boldsymbol{f_{cost} = \underbrace{min}_{maxitr} \frac{1}{n} \sum_{i=1}^n (t_i - \hat{t}_i)^2} \qquad \textbf{(8)}$$

Step.8: for i=1:maxitr,

{

Update weights as:

$$\boldsymbol{w_{i+1} = w_i - \alpha \nabla f_{cost}(w_i) - \left[\begin{bmatrix} \frac{\partial^2 e_1}{\partial w_1^2} & \cdots & \frac{\partial^2 e_1}{\partial w_m^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 e_n}{\partial w_1^2} & \cdots & \frac{\partial^2 e_n}{\partial w_m^2} \end{bmatrix} * \begin{bmatrix} \frac{\partial^2 e_1}{\partial w_1^2} & \cdots & \frac{\partial^2 e_1}{\partial w_m^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 e_n}{\partial w_1^2} & \cdots & \frac{\partial^2 e_n}{\partial w_m^2} \end{bmatrix}^T + \alpha I \right]_i^{-1} * (t_i - \hat{t}_i)} \qquad \textbf{\textit{(9)}}$$

}

Step.9: if ($\boldsymbol{i == maxitr}$ ***or*** $\boldsymbol{f_{cost}}$ ***stabilizes over k-fold, validation)***

{

Truncate training

else

Update weights

}

Step.10: Computer MAPE and $\boldsymbol{R^2}$ ***on convergence.***

Stop.

Here,

The least square optimization is considered as the cost function.

I is an identity matrix.

α is the learning rate.

$t_i\ and\ \hat{t}_i$ are the target and predicted values.

The global influencing factors incorporated through the sentiment polarities have been computed using the Naïve Bayes classifier following the Bayes theorem of conditional polarity given by [37]:

$$\boldsymbol{P}\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)} \qquad \textbf{(10)}$$

Here,

Posterior Probability [P (H/X)] is the probability of occurrence of event H when X has already occurred

Prior Probability [P (H)] is the individual probability of event H

X is termed as the tuple and H is is termed as the hypothesis.

Here, [P (H/X)] denotes the probability of occurrence of event X when H has already occurred.

For overlapping text samples, for tokenization, the class is computed as:

$$f_{class} = \max\left(prob\begin{bmatrix} C_{+ve} \\ C_{-ve} \\ C_{neutral} \end{bmatrix}\right) \qquad \textbf{(11)}$$

Here,

f_{class} is the sentiment class of a feature sample.

The mean absolute error (MAPE:%), root mean squared error (RMSE) and regression (R^2) have been considered as the fundamental evaluation parameters. Each of the parameters has been defined in the subsequent section. The experimental setup for the proposed approach along with the obtained results are presented in the next. A detailed discussion pertaining to the obtained results also ensues.

4. Results And Discussion

The experimental results have been performed on S&P datasets, comprising of daily stock prices of Apple, Google and Tesla. The data used ranges over a period of 10 years, with minimum, maximum, and closing prices being marked. The significant statistical data metrics such as minimum value, maximum value, mean and standard deviation have also been marked on the data. A similar statistical analysis for Google and Tesla stocks over a period of 10 years has been presented subsequently.

Subsequent to the data representation, statistical curve fitting employing both linear and non-linear fitting methods have been shown. Finally the forecasting results without and with regularization has been presented. A similar analysis has been presented for all the 3 stocks. The data has been obtained from the following source:https://finance.yahoo.com/quote/

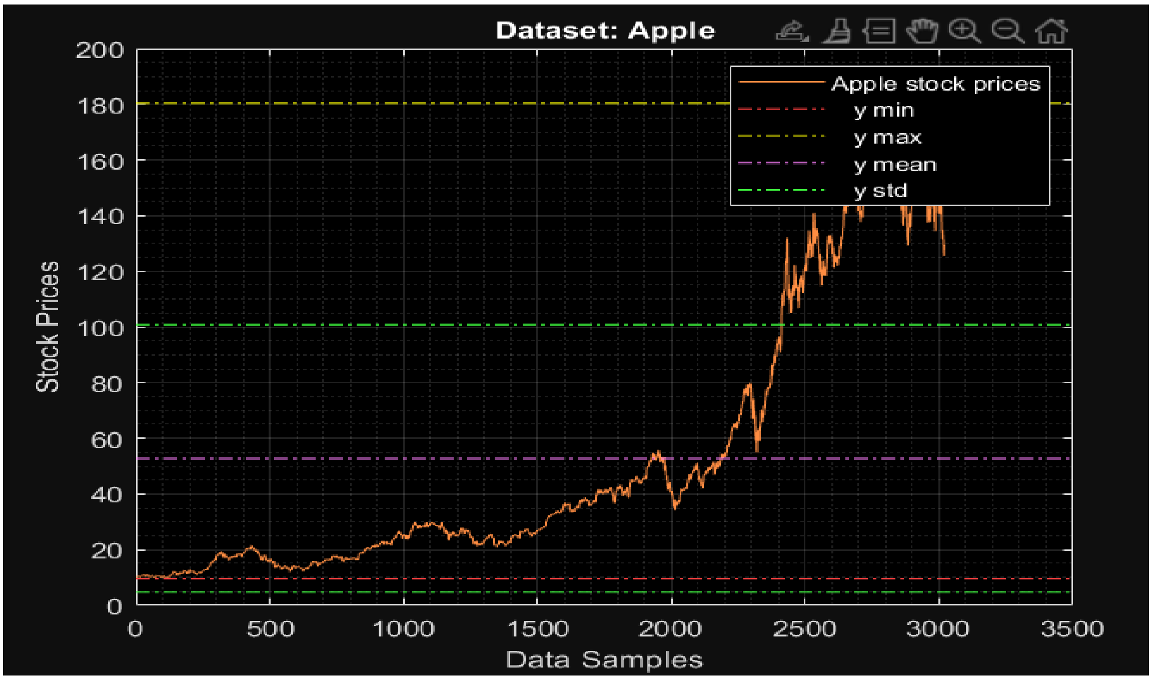

Fig.2 Apple stocks over a 10 year period

Figure 2 depicts the variation of the Apple Stock prices over a period of 10 years.

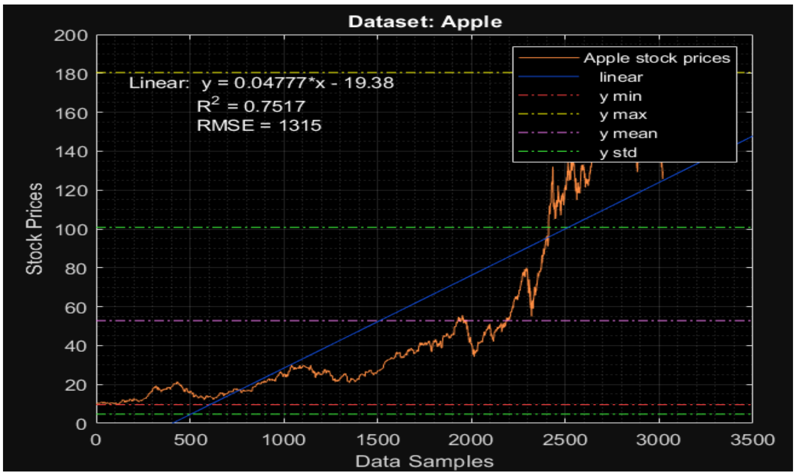

Fig.3 Curve Fitting with Linear Estimation

Figure 3 depicts the curve fitting with linear estimation for Apple stocks.

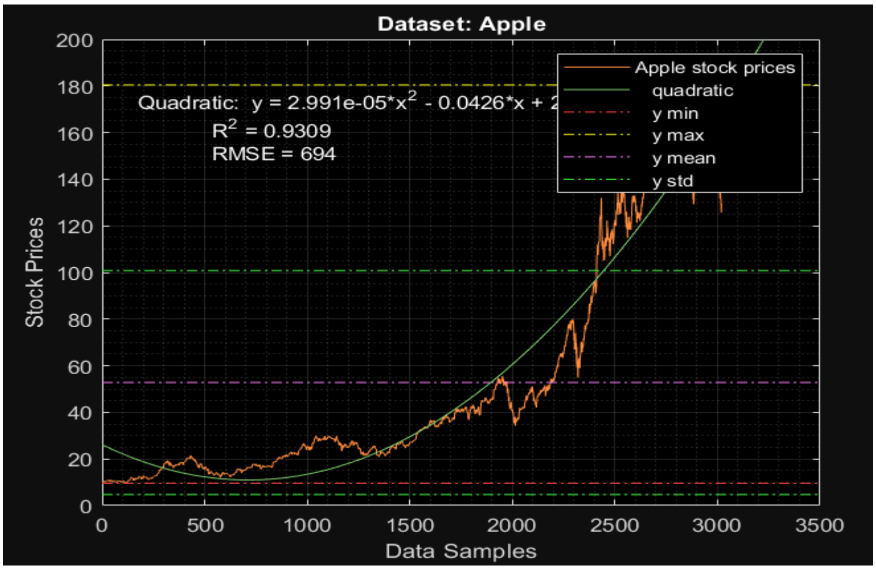

Fig.4 Curve Fitting with Non-Linear Estimation

Figure 4 depicts the curve fitting with non-linear estimation for Apple stocks.

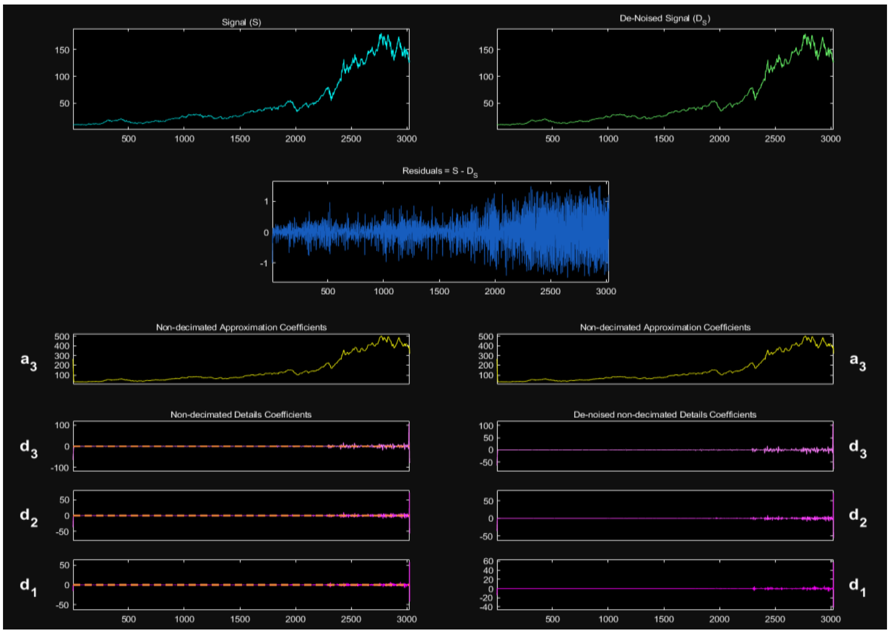

Fig.5 Denoising with stationary Symlet at level 3

Figure 5 depicts the denoising with stationary Symlet for Apple stocks.

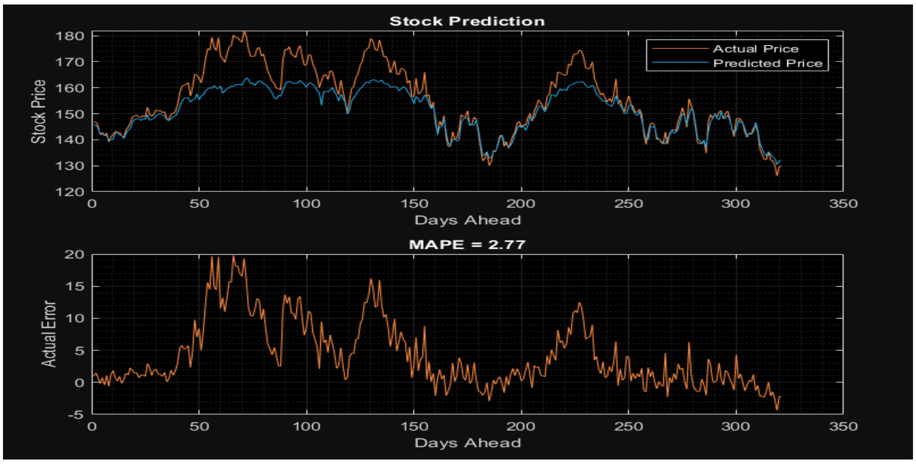

Fig.6 Forecasted and Actual Values

Figure 6 depicts the actual and forecasted stock prices for Apple Stocks, without regularization.

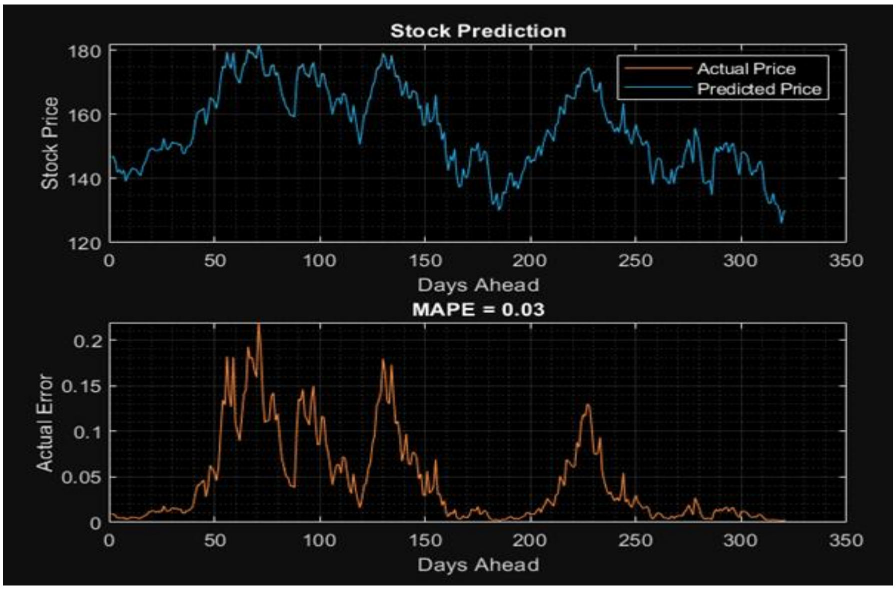

Fig.7 Forecasted and Actual Values with Regularization

Figure 7 depicts the actual and forecasted stock prices for Apple Stocks, with regularization.

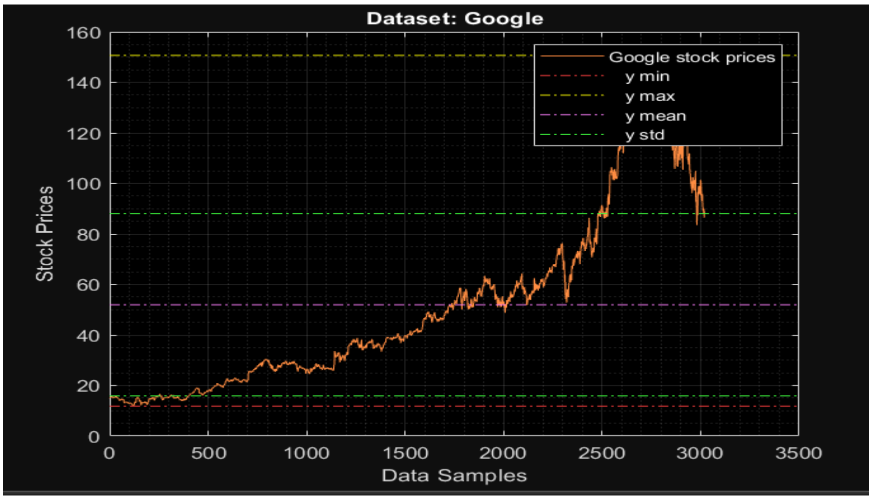

Fig.8 Google stocks over a 10 year period

Figure 8 depicts the variation of the Google Stock prices over a period of 10 years.

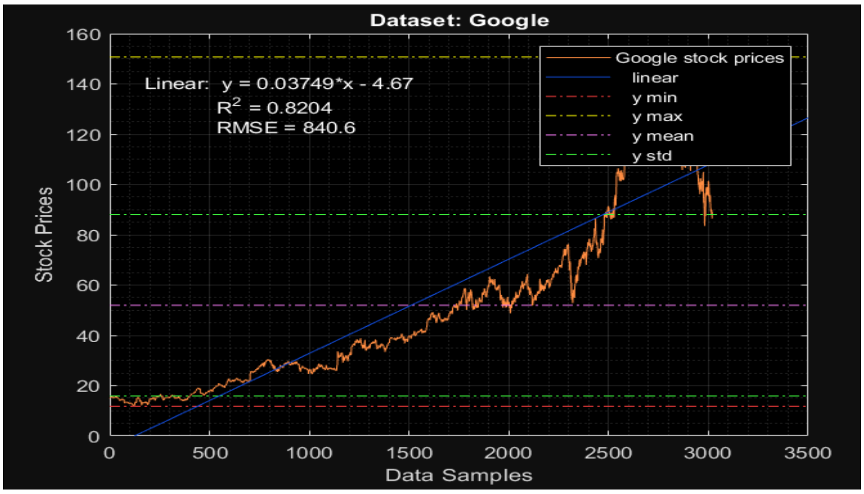

Fig.9 Curve Fitting with Linear Estimation

Figure 9 depicts the curve fitting with linear estimation for Google stocks.

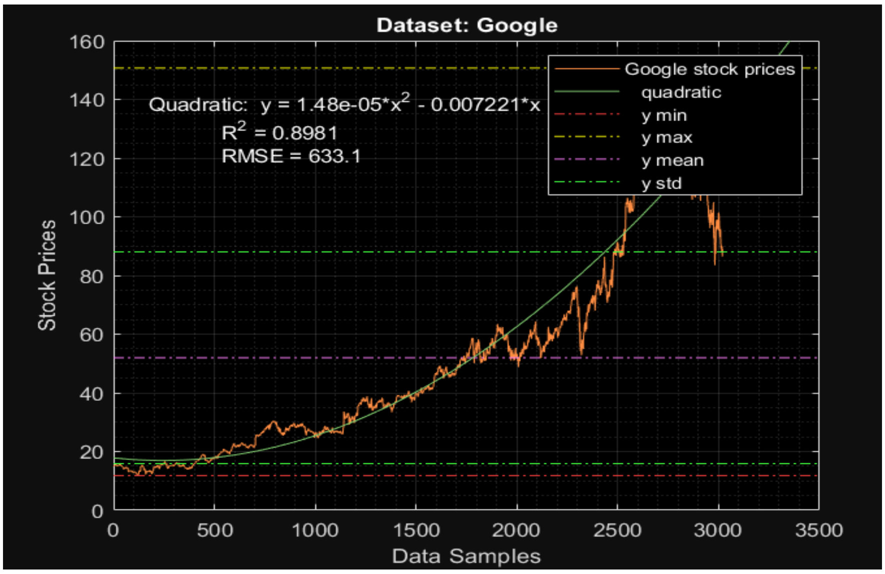

Fig.10Curve Fitting with Non-Linear Estimation

Figure 10 depicts the curve fitting with non-linear estimation for Google stocks.

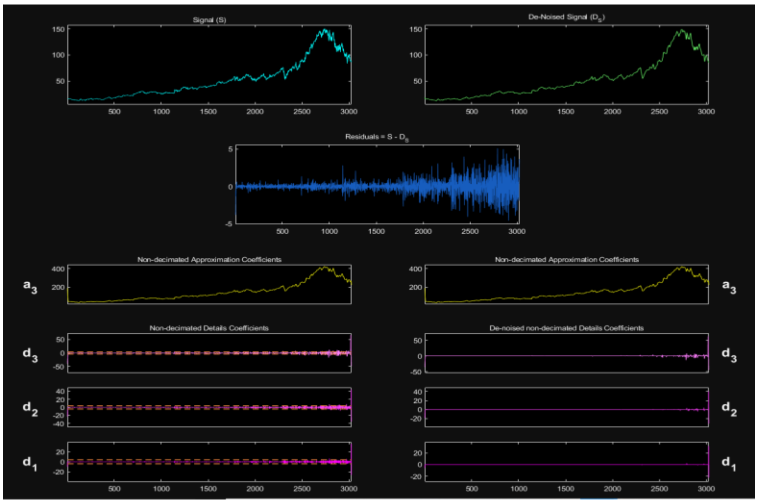

Fig.11 Denoising with stationary Symlet at level 3

Figure 11 depicts the denoising with stationary Symlet for Google stocks.

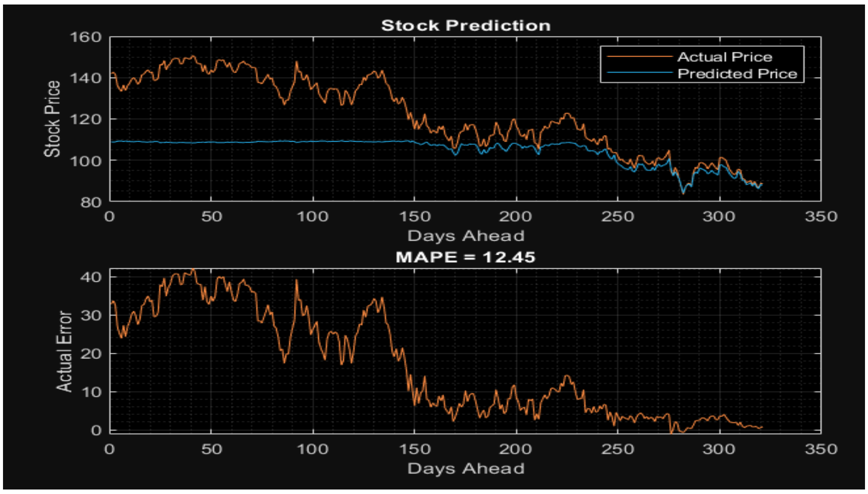

Fig.12 Forecasted and Actual Values

Figure 12 depicts the actual and forecasted stock prices for Google Stocks, without regularization.

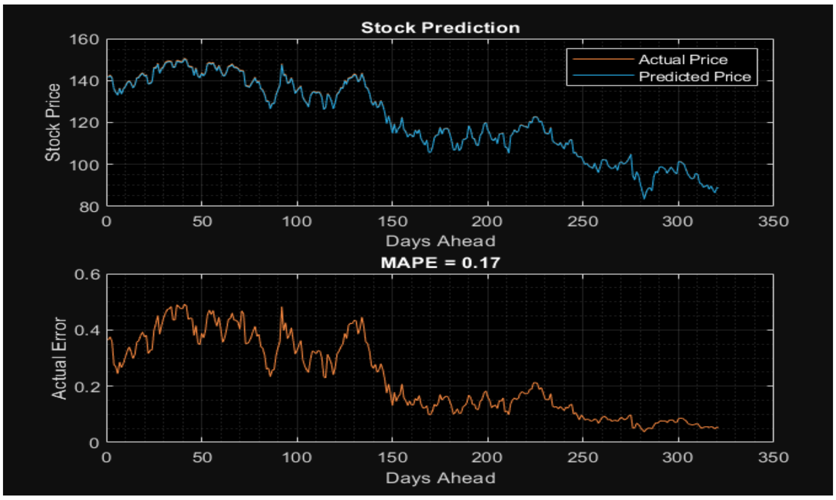

Fig.13 Forecasted and Actual Values with Regularization

Figure 13 depicts the actual and forecasted stock prices for Google Stocks, with regularization.

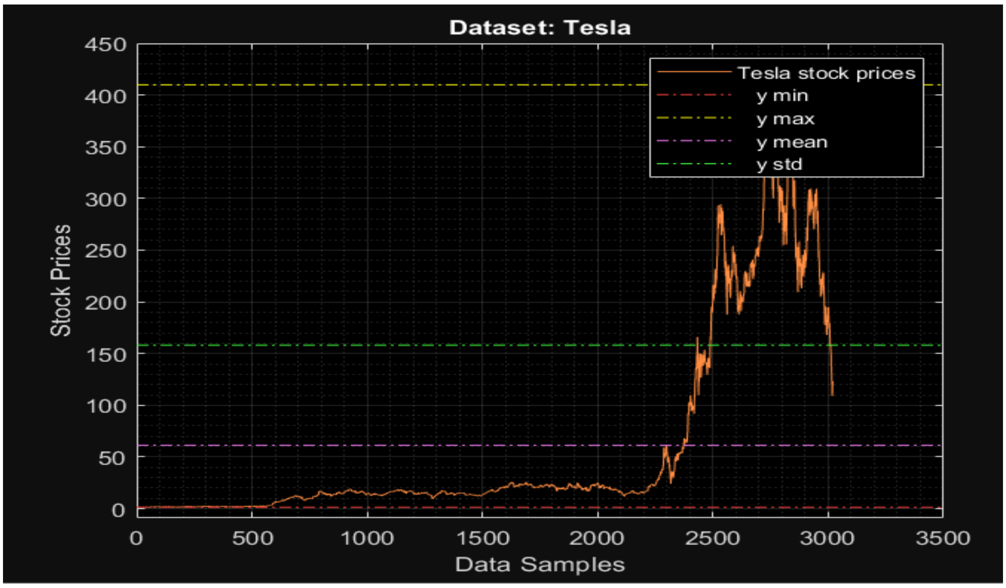

Fig.14 Tesla stocks over a 10 year period

Figure 14 depicts the variation of the Google Stock prices over a period of 10 years.

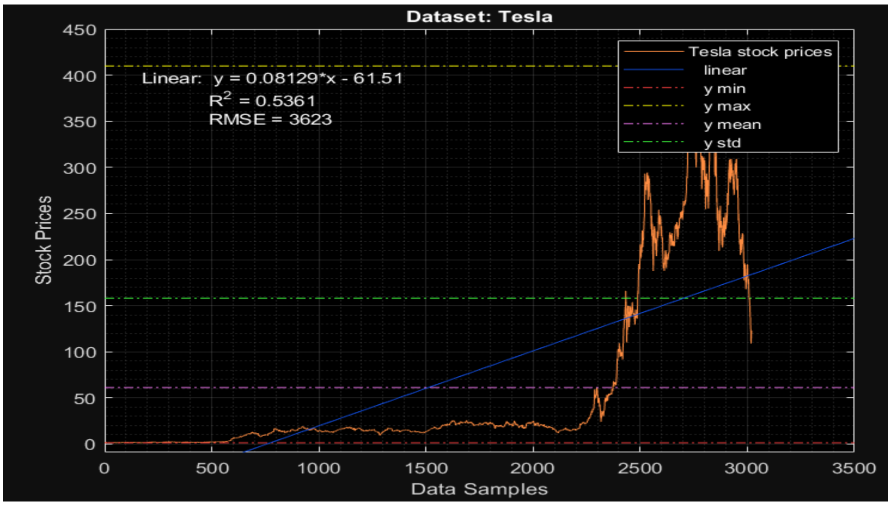

Fig.15 Curve Fitting with Linear Estimation

Figure 15 depicts the curve fitting with linear estimation for Tesla stocks.

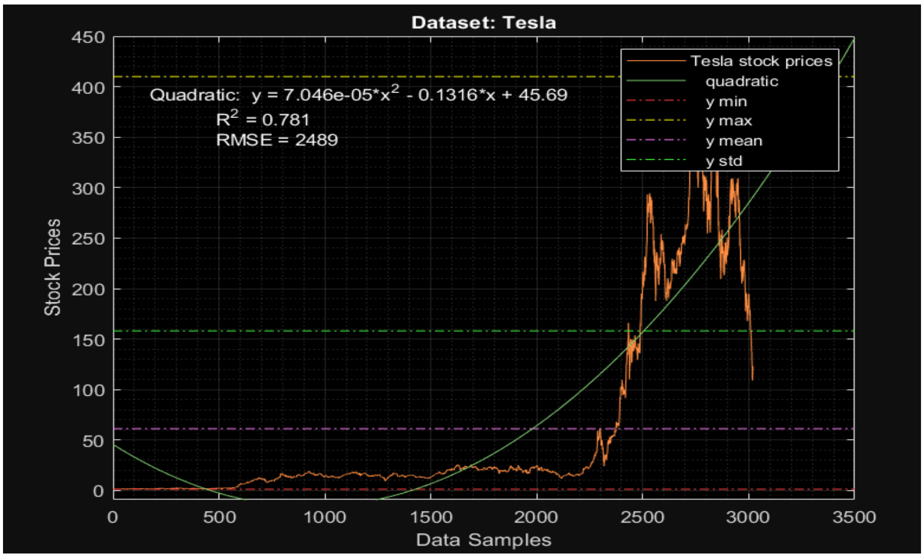

Fig.16 Curve Fitting with Non-Linear Estimation

Figure 16 depicts the curve fitting with non-linear estimation for Tesla stocks.

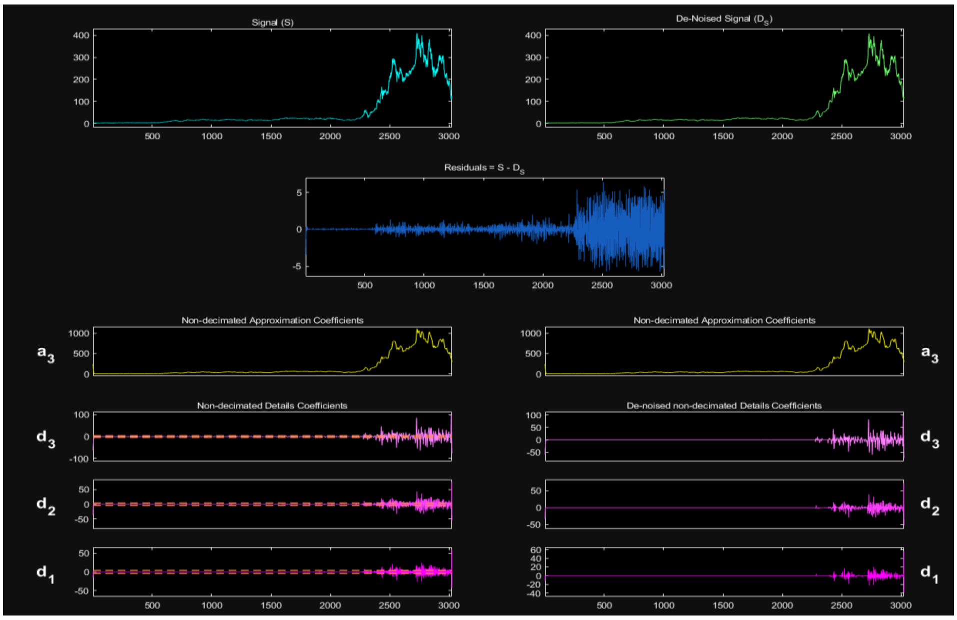

Fig.17 Denoising with stationary Symlet at level 3

Figure 17 depicts the denoising with stationary Symlet for Tesla stocks.

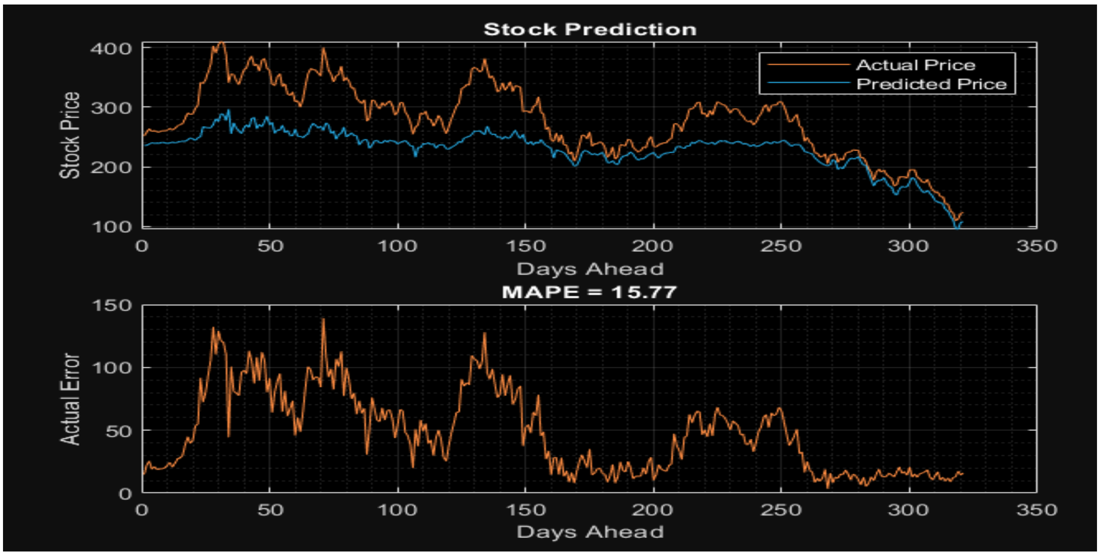

Fig.18 Forecasted and Actual Values

Figure 18 depicts the actual and forecasted stock prices for Tesla Stocks, without regularization.

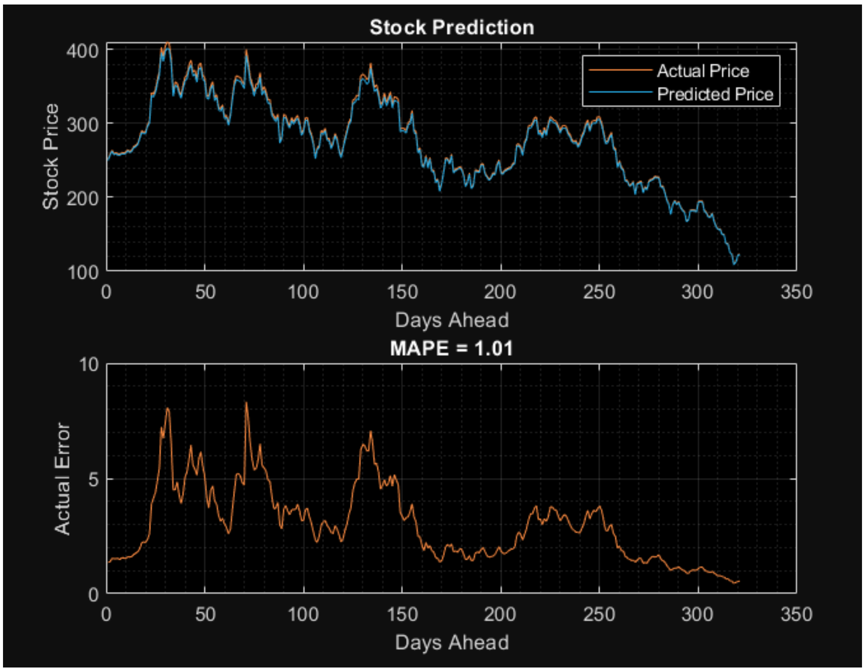

Fig.19 Forecasted and Actual Values with Regularization

Figure 19 depicts the actual and forecasted stock prices for Tesla Stocks, with regularization.

Table 1. Summary of Parameters: Apple

S.No.	Parameter	Value
1.	Denoising	SWT (Sym)
2.	Levels	3
3.	Denoising	Soft
4.	R^2 (Linear)	0.7515
5.	R^2 (Non-Linear)	0.9309
6.	a_0 (Linear)	0.04777

7.	a_1 (Linear)	-19.5
8.	a_0 (Non-Linear)	2.991
9.	a_1 (Non-Linear)	-0.5
10.	a_2 (Non-Linear)	-0.0426
11.	RMSE (Linear)	1315
12.	RMSE (Non-Linear)	694
13.	MAPE (without regularization)	2.77
14.	MAPE (with Regularization)	0.03

Table 2. Summary of Parameters: Google

S.No.	Parameter	Value
1.	Denoising	SWT (Sym)
2.	Levels	3
3.	Denoising	Soft
4.	R^2 (Linear)	0.8204
5.	R^2 (Non-Linear)	0.8981
6.	a_0 (Linear)	0.3749
7.	a_1 (Linear)	-4.67
8.	a_0 (Non-Linear)	1.48
9.	a_1 (Non-Linear)	-5
10.	a_2 (Non-Linear)	-0.007221
11.	RMSE (Linear)	846
12.	RMSE (Non-Linear)	660.1
13.	MAPE (without regularization)	12.45
14.	MAPE (with Regularization)	0.17

Table 3. Summary of Parameters: Tesla

S.No.	Parameter	Value
1.	Denoising	SWT (Sym)
2.	Levels	3
3.	Denoising	Soft
4.	R^2 (Linear)	0.5361
5.	R^2 (Non-Linear)	0.781
6.	a_0 (Linear)	0.08129
7.	a_1 (Linear)	-61.51
8.	a_0 (Non-Linear)	7.046

9.	a_1 (Non-Linear)	-5
10.	a_2 (Non-Linear)	-0.45.69
11.	RMSE (Linear)	3623
12.	RMSE (Non-Linear)	2489
13.	MAPE (without regularization)	15.77
14.	MAPE (with Regularization)	1.01

Table. 4 Comparison with existing techniques

S.No	Authors	Performance
1.	Kim et al.	MAPE: 43%
2.	Wu et al.	MAPE: 50.4%
3.	Sen et al.	MAPE:5.6%
4.	Sharaf et al.	MAPE: 7.7%
5.	**Proposed**	**0.03 (Apple)** **0.17 (Google)** **1.01 (Tesla)**

Tables 1, 2 and 3 summarize all the important stochastic and performance parameters Apple, Google and Tesla datasets. The experiment has been performed on a diverse set of data with Apple being a computed hardware/mobile phone manufacturer, Google being a software company and Tesla being an automobile manufacturer. The analysis is done on a day wise stock dataset ranging over a period of 10 years, from January 2013 to January 2023.The forecasting results have evaluated in terms of the following parameters:

1) Mean Absolute Percentage Error (MAPE), which is defined as:

$$\boldsymbol{MAPE = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{|x_i-\hat{x}_i|}{x_i}\right)} \qquad \textbf{(12)}$$

2) Root Mean Squared Error (RMSE), which is defined as:

$$\boldsymbol{RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\hat{x}_i)^2}{n}}} \qquad \textbf{(13)}$$

3) Regression Value (R^2): The regression value implies the level of similarity among the predicted and actual values.

4) Fitting Co-efficients: The values $(a_0, a_1\ and\ a_2)$ corresponding to the linear or non-learn fitting relation given by:

$$\boldsymbol{y = a_0 + a_1x + a_2x^2 + \cdots \ldots\ldots a_nx^n} \quad \textbf{(14)}$$

The first step of analysis for each of the data sets is the nature of fitting for the datasets. Linear and Non-Linear (Quadratic) curve fitting has been performed to compute the $\boldsymbol{R^2\ and\ RMSE}$ values of the datasets. It has been observed that the non-linear $\boldsymbol{R^2}$ value in higher compared to that for linear fitting indicating the non-linearity inherently present in the datasets. The$\boldsymbol{RMSE}$ value though reduces indicating better fitting results for the non-linear model. The above inferences could have also been intuitively garnered, however prior to design of a model, the nature of the datasets were verified, so as to opt for optimal design of the machine learning model.

The fitting co-efficient values $(\boldsymbol{a_0\ and\ a_1})$for the linear model and $(\boldsymbol{a_0, a_1\ and\ a_2})$ for the non-linear models have been presented. Subsequently, the wavelet transform (stationary) with Sym (Symlet) at level 3 has been performed to de-noise the data. The wide sense stationarity is an attribute which is typically observed across time series datasets (indicating seasonality). The low and high frequency co-efficients $(\boldsymbol{C_a\ and\ C_d})$values for the non-decimated and decimated waveforms for both original and de-noised data have been computed. Increase in the number of levels clearly indicate the reduction in the noise amplitude for all the datasets. Separation of the data as $\boldsymbol{Residual = S + d_3}$has also been shown. Truncating the decomposition has been done at level 3 as increasing the levels further typically increases the computational complexity, introduces decimation and doesn't result in significant noise removal.

Finally, the forecasting has been performed for 2 separated cases, one being without the regularization parameter and the other being with the regularization parameter. An important observation which can be made is the fact that the proposed weighted mean based attention model achieves relatively low MAPE for the Apple dataset, being 2.77 (without regularization) but clearly results in much higher MAPE for the Google and Tesla datasets (being 12.47 and 15.77) respectively. This observation is an indication of the data specific performance of machine learning and deep learning models which show large divergences across datasets. To standardize the model across a multitude of datasets, the regularization parameter has been incorporated which is essentially a penalty imposed on the training algorithm in case the change in weights result in sudden change in the monotonicity pattern of the cost function. The penalty based regularization metric can be computed as:

$$\boldsymbol{r_{cost}} = \frac{|\boldsymbol{w}|}{\boldsymbol{J}} \nabla \boldsymbol{w}; \boldsymbol{i} = \boldsymbol{1}: \boldsymbol{n} (\mathbf{12})$$

Here,

r_{cost} denotes the regularization factor.

$|w|$ denotes the weight vector.

J denotes the cost function.

i denotes iterations

The introduction of the regularization parameter clearly reduces the MAPE manifolds, clearly exhibiting the effectiveness of the regularization been employed.

Table 4 compares the performance of the proposed system against contemporary research in the domain. The comparison yields the superior performance of the proposed approach again the existing approaches.

Conclusion

This paper presents a weighted attention model for stock market prediction incorporating the impacts of global influencing features. Sentiment polarity has been added as an exogenous data metric through opinion mining, to create a composite training vector. The attention based model tried to incorporate both the impacts of seasonality and local randomness over shorter intervals of time series analysis. Denoising has been performed using the stationary Symlet to filter raw data. A penalty based regularization parameter has also been included to reduce the MAPE of the proposed system. The forecasting MAPE has been computed with and without the regularization parameters whose results clearly show the reduction in MAPE through the use of the penalty based regulation. The choice of the datasets has been kept diverse to validate the effectiveness of the proposed model. A clear illustration of the data specific nature of machine and deep learning models has been augmented through the divergence in MAPE across datasets, thereby reinforcing the need for regularization. The final MAPE values attained are 0.03, 0.17 and 1.01 for Apple, Google and Tesla datasets respectively. A comparative analysis with contemporary approaches supports the improved performance of the proposed model.

References

[1] A. W. Li and G. S. Bastos, "Stock Market Forecasting Using Deep Learning and Technical Analysis: A Systematic Review," in IEEE Access, vol. 8, pp. 185232-185242.

[2] P He, Y Sun, Y Zhang, T Li, “COVID–19's impact on stock prices across different sectors—An event study based on the Chinese stock market”, Emerging Markets Finance and Trade, Taylor & Francis, 20202, vol.56., no.10, pp. 2198-2212.

[3] E Blankespoor, E deHaan, I Marinovic, “Disclosure processing costs, investors' information choice, and equity market outcomes: A review”, Accounting and Economics, Elsevier, 2020, vol.70., no.2., art.no. 101344.

[4] E Chong, C Han, FC Park, “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies”, Expert Systems with Applications, Elsevier, 2017, vol.83., pp. 187-205.

[5] S. Singh, T. K. Madan, J. Kumar and A. K. Singh, "Stock Market Forecasting using Machine Learning: Today and Tomorrow," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 738-745.

[6] M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in IEEE Access, 2020, vol. 8, pp. 150199-150212

[7] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis," 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 2019, pp. 205-208.

[8] AM Rather, VN Sastry, A Agarwal, Stock market prediction and Portfolio selection models: a survey”, OPENRESEARCH, Springer, 2017., vol.54, pp.558–579.

[9] A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2017, pp. 506-509.

[10] B Li, KCC Chan, C Ou, S Ruifeng, “Discovering public sentiment in social media for predicting stock movement of publicly listed companies”, Information Systems, 2017, vol.69., pp. 81-92.

[11] X Zhang, Y Zhang, S Wang, Y Yao, B Fang, “Improving stock market prediction via heterogeneous information fusion”, Knowledge Based Systems, Elsevier 2018, vol.143., pp. 236-247.

[12] S Kim, S Ku, W Chang, JW Song, “Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques”, IEEE Access 2020, Vol-8, pp: 111660 – 111682.

[13] X Li, P Wu, W Wang, “Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong”, Information Processing & Management, Elsevier 2020. Volume 57, Issue 5, pp: 1-19.

[14] J. Sen, S. Mehtab and A. Dutta, "Volatility Modeling of Stocks from Selected Sectors of the Indian Economy Using GARCH," 2021 Asian Conference on Innovation in Technology (ASIANCON), IEEE, 2021, pp. 1-9.

[15] D Kobiela, D Krefta, W Król, P Weichbroth, “ARIMA vs LSTM on NASDAQ stock exchange data”, Procedia in Computer Science, Elsevier 2022, vol.207, pp.3830-3839.

[16] M Sharaf, EED Hemdan, A El-Sayed, NA El-Bahnasawy, “An efficient hybrid stock trend prediction system during COVID-19 pandemic based on stacked-LSTM and news sentiment analysis”, Multimedia Tools and Applications, Springer 2023, vol. 82, pp.23945–23977.

[17] G. Li, A. Zhang, Q. Zhang, D. Wu and C. Zhan, "Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction," in IEEE Transactions on Circuits and Systems II: Express Briefs, 2022, vol. 69, no. 5, pp. 2413-2417.

[18] G. Duan, M. Lin, H. Wang and Z. Xu, "Deep Neural Networks for Stock Price Prediction," 2022 14th International Conference on Computer Research and Development (ICCRD), Shenzhen, China, 2022, pp. 65-68.

[19] R. Ren, D. D. Wu and T. Liu, "Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine," in IEEE Systems Journal, vol. 13, no. 1, pp. 760-770.

[20] F. Gers, D. Eck, J. Schmidhuber,“Applying LSTM to time series predictable through time-window approaches” Neural Nets WIRN Vietri-01. Perspectives in Neural Computing. Springer, 2020, pp.193-200.

[21] X. Luo, W. Qin, A. Dong, K. Sedraoui, M. Zhou, “Efficient and High-quality Recommendations via Momentum-incorporated Parallel Stochastic Gradient Descent-Based Learning”, IEEE/CAA Journal of Automatica Sinica 2021, vol.8, no.2, pp.402-411.

[22] D. Shah, H. Isah and F. Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4705-4708.

[23] TM Nisar, M Yeung, “Twitter as a tool for forecasting stock market movements: A short-window event study”, Journal of Finance and Data Science, Elsevier 2018, vol.4., no.2., pp. 101-119.

[24] M. F. R. Abu Bakar, N. Idris, L. Shuib and N. Khamis, "Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work," in IEEE Access, 2020, vol. 8, pp. 24687-24696

[25] V. Kumar and B. Subba, "A TfidfVectorizer and SVM based sentiment analysis framework for text data corpus," 2020 National Conference on Communications (NCC), Kharagpur, India, 2020, pp. 1-6.

[26] G Kumar, S Jain, UP Singh, “Stock market forecasting using computational intelligence: A survey”, Archives of Computational Methods in Engineering, Springer, 2021, vol.28, pp. pages1069–110.

[27] M. Wen, P. Li, L. Zhang and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series," in IEEE Access, vol. 7, pp. 28299-28308.

[28] S Patel, BDV Surya, C Manjunath, B Marimuthu, “Forecasting Stock Market Indexes Through Machine Learning Using Technical Analysis Indicators and DWT”, Congress on Intelligent Systems, LNDECT Book Series, Springer 2022, pp. 625–638.

[29] S. Saxena, R. Jais and M. K. Hota, "Removal of Powerline Interference from ECG Signal using FIR, IIR, DWT and NLMS Adaptive Filter," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0012-0016.

[30] K. Fukumori, H. T. Thu Nguyen, N. Yoshida and T. Tanaka, "Fully Data-driven Convolutional Filters with Deep Learning Models for Epileptic Spike Detection," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 2772-2776.

[31] G. Liu, F. Xiao, C. -T. Lin and Z. Cao, "A Fuzzy Interval Time-Series Energy and Financial Forecasting Model Using Network-Based Multiple Time-Frequency Spaces and the Induced-Ordered

Weighted Averaging Aggregation Operation," in IEEE Transactions on Fuzzy Systems, 2020, vol. 28, no. 11, pp. 2677-2690.

[32] Z. Wang, W. Yan and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 2017, pp. 1578-1585.

[33] M. Zhang, X. Li and L. Wang, "An Adaptive Outlier Detection and Processing Approach Towards Time Series Sensor Data," in IEEE Access, vol. 7, pp. 175192-175212, 2019.

[34] X. Wang and C. Wang, "Time Series Data Cleaning: A Survey," in IEEE Access, 2020, vol. 8, pp. 1866-1881.

[35] P Netrapalli, "Stochastic gradient descent and its variants in machine learning", Journal of the Indian Institute of Science, Springer 2019, vol.99, no.2., pp.201-2013.

[36] W Du, D Côté, Y Liu -, "Expert Systems with Applications, Elsevier, 2023, vol.219., pp. 119619.

[37] Y. Shen, M. Chen, S. Cai, S. Hu and X. Chen, "BERT-RF Fusion Model Based on Bayesian optimization for Sentiment Classification," 2022 IEEE 2nd International Conference on Computer Communication and Artificial Intelligence (CCAI), Beijing, China, 2022, pp. 71-75.