Stock Analysis Based on Trends in Crude Oil Prices

Prudhvi Raj Patnala, Nishant Ravi Shankar, Abhishikta Sai and Gowri Srinivasa

November 28, 2021

Stock Analysis Based on Trends in Crude Oil Prices

Patnala Prudhvi Raj*, Nishant Ravi Shankar†, P. Abhishikta Sai‡ and Gowri Srinivasa§
PES Center for Pattern Recognition, PES University EC Campus,
Department of Computer Science and Engineering, PES University RR Campus, Bengaluru, India.
Email: {*pruthvipatnala, †nishantravishankar, ‡abhishikta.sai21}@gmail.com, §gsrinivasa@pes.edu

***Abstract*—Analyzing stock prices is a key component of financial analysis of a company. Traditional stock price forecasting methods model historical price data by considering Open, High, Low, Close (OHLC) values to analyze performance of companies across various sectors such as oil refineries and logistics, for example. This approach likely misses the impact of certain critical driving factors, such as crude oil, on the company, whereas in practice, crude oil would play a major role in the performance of oil-refineries and logistics companies. This paper focuses on the inclusion of crude oil data in the analysis and its impact on forecasting. Patterns in the trends of oil prices would also have an impact on the performance of various other industries. This study would particularly help long term investors understand how the performance trends of a company would fluctuate based on the industry it belongs to. We present a few insights on analyzing company stock price trends for different industries as a function of oil price trends.**

***Index Terms*—Keywords - Crude Oil, Multivariate Forecasting, ARIMAX, SVM, LSTM, Trend Analysis.**

I. Introduction

Crude Oil is a primary energy source for India, and ranks second after coal. India imports around 70% of its crude oil, and hence an increase in the prices of crude oil is generally thought to increase inflation, causing a major slowdown in economic growth [4]. An increase in the price of crude oil also translates to an increase in the cost to meet domestic fuel needs and an increase in operational costs of oil-dependent companies. We found crude oil to be a critical driving factor that both upstream and downstream industries depend on for generating revenue. Forecasting of stock prices has been carried out by considering oil refining companies such as Indian Oil Corporation (IOC) Limited, Oil and Natural Gas Corporation (ONGC) Limited and logistics companies such as Transport Corporation of India (TCI) Limited.

It has been observed that the fall in oil prices is good for the rupee, but does not affect the Indian current account deficit due to strengthening of the US dollar at the same time [9]. Therefore, both the rise of international crude oil prices, and the import dependency cause damage to the Indian Economy. Oil refining companies, on the other hand, get negatively affected due to the fall in oil prices, mainly because their stock prices rely on the exports of the refined crude oil.

The rest of the paper has been organized as follows - details of related work found in literature is presented in Section II, followed by the methodology adopted in Section III with details of data collection and pre-processing, models designed and the approach for trend analysis and forecasting, followed by experimental results and discussions in Section IV and finally, a Section on Conclusions that presents some insights gleaned through this work.

II. Literature Survey

Machine learning techniques mainly work on homogeneous, stationary data and perform a ”one-shot” analysis of the data. This is not particularly useful when it comes to price forecasting, since new data is being generated everyday and in order to predict the future prices, the entire data-set will have to be used again to forecast.

The relationship between fluctuation of crude oil prices and changes in the Indian stock market has been investigated using trend analysis, correlation analysis and regression modeling on ten year data of crude oil price (Brent Crude) and India Stock Market Index (BSE) [7]. The findings suggest a weak but significant relation between changes in crude oil price in the international markets due to various socio-economic turmoils and returns in Indian stock market.

Another study has focused on the long term and short term impact of change in crude oil prices on Gulf Corporation Council (GCC) stock markets using auto-regressive distributive lag (ARDL) to analyze monthly oil prices data and stock market over a ten year period [1]. The results of this study suggests that there is no long term correlation between stock market and crude oil price except in case of Oman. In the short term model, crude oil price has a strong and positive relationship with the change in stock market price in all the GCC countries but the significance of impact varies from one country to another.

A study of the volatility of crude oil prices and its impact on the Indian stock market using data from the Power database from 2006 to 2015 was modeled using GARCH(1,1) model to measure the volatility of Sensex and linear regression was used to find fluctuations in crude oil prices [8]. Their findings suggest a strong impact of crude oil prices on Sensex and was competent to transmit shocks to Sensex.

A recent study has classified the forecasting of crude oil prices into three approaches (1) heuristic approaches (2) econometric models (3) machine learning techniques [3]. While most of the previous work on crude oil price determination was based on forecasting the price within a supply and demand framework, Support Vector Machines (SVMs) has also been used for crude oil time series prediction [11].

We seek to use the insights from these previous approaches on determining crude oil prices to forecast stock prices of oil-dependent companies using oil-related exogenous variables.

III. Methodology

The first step towards analyzing stock prices of oil-related companies in conjunction with crude oil price fluctuation involves collecting and pre-processing the data required for modeling and forecasting. Once a model is trained based on available data, these models are used to forecast stock prices on test-data, followed by the performance analysis of these models using three measures - Root Mean Squared Error (RMSE), Co-Integration Factor and an Algorithmic Trading function.

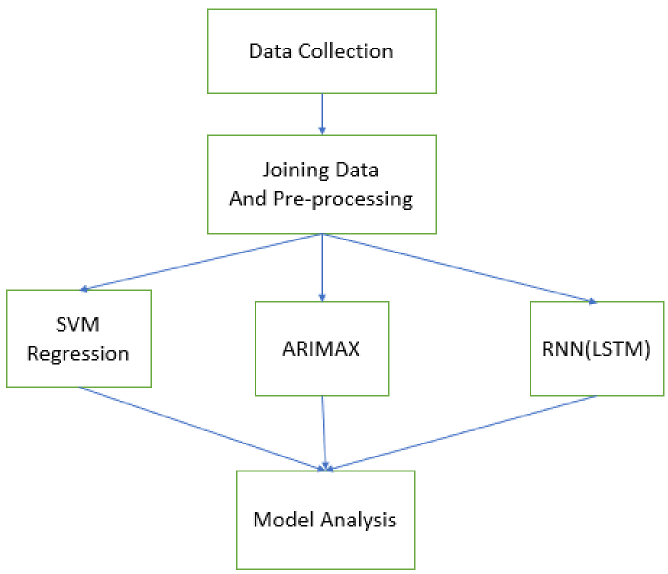

Fig. 1. A flow chart of the solution approach.

A. Data Collection and Pre-processing

The stock prices of different companies have been obtained from the Bombay Stock Exchange (BSE) online portal. The crude oil commodity price history has been collected from the "investing.com" website. The constraints faced are mainly in finding a substantial number of data-points for the data-set. For stock prices, there are a lot of missing dates. For these, besides the weekends, an assumption that the missing dates are due to the stock market being closed during festivals has been made. Finding matching frequencies of data-points is also a major challenge. For the daily data, since there is a small number of mismatched dates, an intersection of the dates has been taken to avoid unnecessary interpolations and null value generations. Import statistics of only monthly data is freely available, and data-interpolation of that monthly data is avoided as the number of interpolated data-points can exceed the total number of actual import statistics, thus adding no value and increasing noise in the data. For trend analysis, industries that use crude oil as a core component for their operations and those that are dependent on one of its by-products have been considered. Each active company listed in BSE under each of the industries was considered for analyzing how that industry was being affected by crude oil.

B. Models

In this Section, we explore modeling the data we have collected using multiple models.

1) SVM Regression: Support Vector Machines (SVMs) are popular models for classification problems. The underlying principle can also be applied to regression tasks in the form of Support Vector Regression (SVR) [6]. ONGC, IOC and TCI stock prices have been forecast using SVR. The kernel trick, which is used to map non-linearly separable data instances into a higher dimensional space to achieve separability, has been applied to this model. Radial Bias Function Kernel with various gamma values has been applied to perform forecasting. Fig. 2 shows a graph of the actual versus predicted values of closing prices for IOC. We can compare this with the actual versus predicted values for IOC using OHLC with crude oil (see Fig. 3). We notice the model is more accurate with crude oil being factored into the analysis. However, we explore other options to see if the model can be further improved.

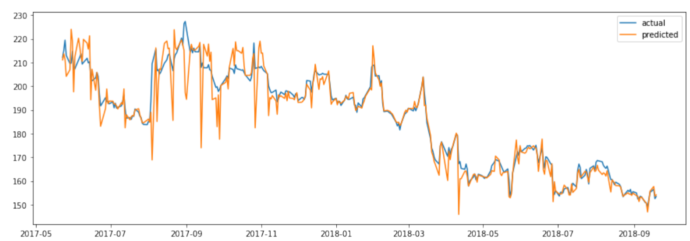

Fig. 2. SVR - Actual vs Predicted for Closing Prices of IOC Using OHLC only

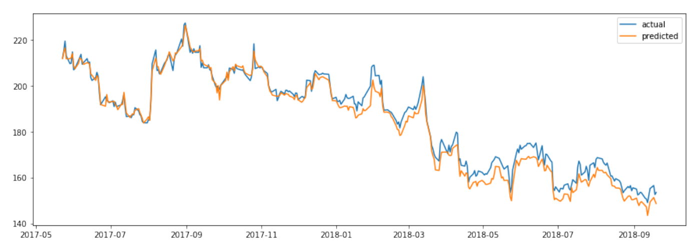

Fig. 3. SVR - Actual vs Predicted for Closing Prices of IOC using OHLC with Crude Oil

2) ARIMAX: Since stock prices are strongly related to past data, we seek to include that in the model. The Auto-regressive Integrated Moving Average (ARIMA) is a generalized version of Auto-regressive moving average (ARMA) model used for for analysis and forecasting of time series data [2], [10]. ARIMA with an explanatory variable (ARIMAX), an extension of the ARIMA model, can include exogenous variables with the auto-regressive and moving average variables for forecasting. The endogenous variables in the ARIMAX (p,d,q) model are the past values of the closing stock prices of each of the oil dependent companies. There were two approaches based on the choice of the exogenous variables (X) variables: (i) OHCP along with the crude oil parameters and (ii) OHCP without crude oil parameters.

The standard tests have been conducted for the choice of (p,d,q) parameters in the ARIMAX model. Through the Augmented Dickey-Fuller (ADF) test, it has been found that most of variables are non stationary, with 'p' value greater than 0.05. To achieve stationarity, a differencing of 1 has been done. Then, the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots have been used to get the appropriate 'p' and 'q' parameters. Along with the spikes in ACF and PACF, the AIC (Akaike information criterion) and the BIC (Bayesian information criterion (BIC) of the first 6 lags have been calculated to get the 'p' and 'q' parameters (see Fig. 4 and Fig. 5). The spikes at higher lags have been ignored since they may be spurious and also since a significant computational power is required to calculate the AR for such a lag. The ACF and PACF for each stock does not show any seasonality, and hence a seasonal ARIMAX (SARIMAX) model has not been resorted to.

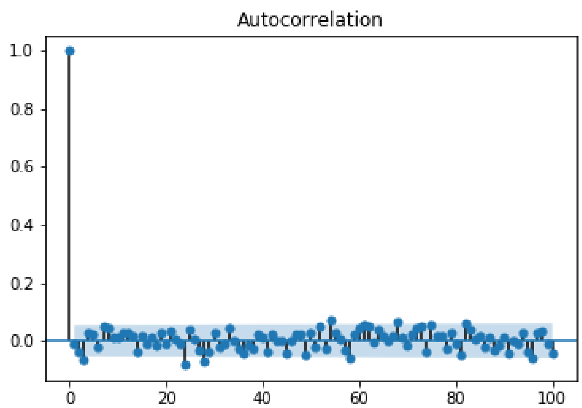

Fig. 4. Auto-Correlation Function For IOC Stock Prices

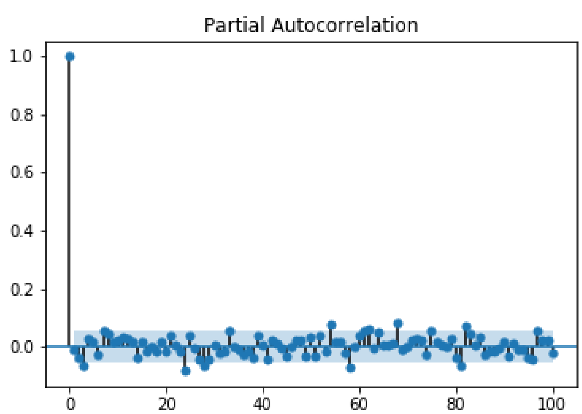

Fig. 5. Partial Auto-Correlation Function for IOC Stock Prices

ARIMAX(p,0,q) model has then been applied on the out-of-sample, non-stationary Y and exogenous values too for forecasting the stock prices of each company, which is after selecting the suitable 'p' and 'q' values. Fig. 6 shows a graph of the actual versus predicted values for closing prices of IOC with crude oil and Fig. 7 shows a plot of the actual versus predicted values for the closing price without crude oil. We note that both graphs seem more accurate than their SVR counterparts shown in Fig. 2 and Fig. 3 respectively.

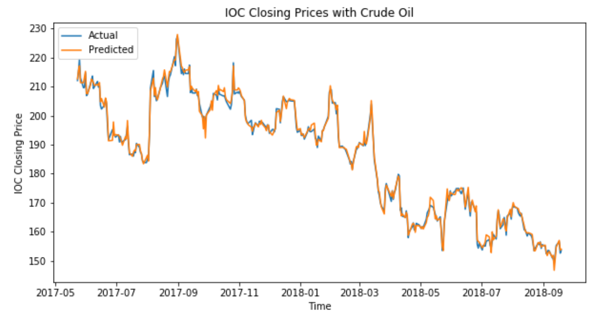

Fig. 6. ARIMAX - Actual vs Predicted for Closing Prices of IOC using OHLC with Crude Oil

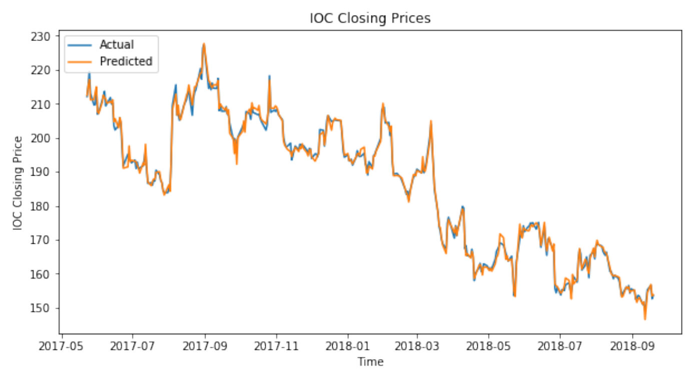

Fig. 7. ARIMAX - Actual vs Predicted for Closing Prices of IOC using OHLC without Crude Oil

3) LSTM: A popular approach to forecasting in the recent times has been the use of Long Short-Term Memory (LSTM) networks. LSTM have the characteristic of learning to store selective information and patterns over longer periods of time [5]. A typical LSTM unit has a cell, an input gate, an output gate and a forget gate. The purpose of the three gates is to control the information flow into and out of the cell. Prior to implementation of the model, the data has been normalized to fit to a supervised learning model. A sequential model with 50 LSTM cells has been implemented with an Adam optimizer. A graph of the actual versus predicted values of closing prices of IOC using LSTM is shown in Fig. 9, whereas Fig. 8 shows the predicted values of IOC taking into account crude oil prices.

A visual inspection of the graph indicates the predicted values are smoother (less stationarities) with the LSTM whereas the ARIMAX model seems to be more sensitive to local changes.

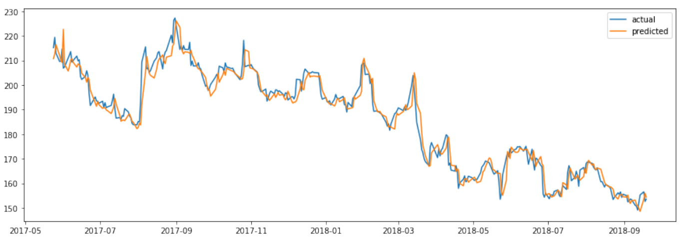

Fig. 8. LSTM - Actual vs Predicted for Closing Prices of IOC with Crude oil

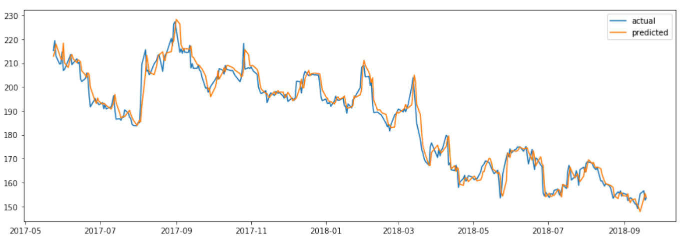

Fig. 9. LSTM - Actual vs Predicted for Closing Prices of IOC

C. Quantitative Evaluation of Forecasting Algorithms

- *RMSE* - The root mean squared error captures the error in predictions with respect to the actual values. This is an indication of how good a forecasting model is.
- *Co-integration* - This metric captures how well two series fluctuate together rather than how close the predicted values are. A hypothesis test is conducted to check the co-intergration of two series. The Null hypothesis is that there is no co-integration while the alternative hypothesis is that there is a Co-integrating relationship. If the $p - value$ is small, below a critical size, then we can reject the null hypothesis.
- *Algorithmic Trading* - An algorithmic trading model was used as one of the metrics for evaluating the implemented forecasting models. The trading model was designed for simulating stock exchanges on a daily basis. The model uses the stock prices forecast for simulating stock transactions. It follows the idea of selling stocks if the stock price forecast for the next day is lower and buying stocks otherwise. The actual stock price values over a period along with its predicted values are taken as input. Stock trades are simulated over the given period using the forecast values and the sequence of buy and sell actions are noted. As a prescription to the investor, this sequence of actions is followed using the actual stock price values to realize the actual net profit. Algorithm 1 presents the steps of Algorithmic Trading.

D. Trend Analysis

We studied 17 different industries (both upstream and downstream) to see how trends in these industries vary with crude oil trends. The trend component of a time series was extracted by analyzing the plot of the time series. A high correlation of the time series with a straight line joining the start and end points of the time series helped determine trend. Analysis was carried out on a yearly, half-yearly, quarterly and monthly basis. A total time period of seven years (2012 - 2018) was considered for analysis. The trend in the industry is swayed by how most of the companies in that industry are affected when crude oil trend shifts. So the industry trend for a time period can be assumed as the most frequent trend across the companies in an industry. The crude oil trend over a time period and the corresponding most frequent trend of that industry was considered for analysis.

Finding the correlation involved hypothesizing what happened when crude oil trend goes up/ down or demonstrates

Algorithm 1 Algorithmic Trading

procedure TRADESIMULATION($forecast$)
$money \leftarrow 0$
$haveStock \leftarrow 1$ // $\#stocks\ currently\ available$
$buyIndices \leftarrow [\ ]$
$sellIndices \leftarrow [\ ]$
$n \leftarrow$ length($forecast$)
for $day \leftarrow 2$ to n **do**
  // $priceDay = forecast[day]$
  if $priceTomorrow > priceToday$ and $money > priceToday$ **then**
    // $buying\ the\ stock$
    $\#stocksToBuy \leftarrow money/priceToday$
    $add\ (today,\ \#stocksToBuy)\ to\ buyIndices$
    $money \leftarrow money - \#stocksToBuy * priceToday$
    $haveStock \leftarrow haveStock + \#stocksToBuy$
  else if $priceTomorrow < priceToday$ and $haveStock \neq 0$ **then**
    // $selling\ the\ stock$
    $add\ (today, haveStock)\ to\ sellIndices$
    $money \leftarrow money + haveStock * priceToday$
    $haveStock \leftarrow 0$
  end if
end for
$netProfit \leftarrow money + haveStock * priceLastDay - 1 * priceFirstDay$
return $netProfit,\ buyIndices,\ sellIndices,\ haveStock$

procedure FOLLOWSIMULATION(actual, forecast)
$expected\ (Profit,\ buyIndices,\ sellIndices,\ HaveStock) \leftarrow TRADESIMULATION(forecast)$
$money \leftarrow 0$
$haveStock \leftarrow 1$
for $(buyday, \#stocksToBuy)\ in\ buyIndices$ **do**
  $money \leftarrow money - actual[buyday] * \#stocksToBuy$
end for
for $(sellday, haveStock)\ in\ sellIndices$ **do**
  $money \leftarrow money + actual[sellday] * haveStock$
end for
$actualNetProfit \leftarrow money + haveStock * actual[lastDay] - actual[firstDay]$
return $actualNetProfit$

no trend. To compute the correlation score, we look at the percentage of the time series for which the actual trend follows our hypothesis. If the score is greater than a threshold, we have an insight. The threshold was empirically set to 70% during our analysis. Such analysis would help investors predict the trend of a company given its industry and crude oil trend.

IV. Experimental Results

In this section, we discuss the experimental results of forecasting and trend analysis. Forecasting of stock prices focused on three companies that are influenced by crude oil price fluctuations. Companies under various industries influenced by crude oil have been considered for trend analysis.

A. Forecasting

Table I and Table II show the results of forecasting of stock prices using the three evaluation measures mentioned above, for OHLC + crude oil as exogenous variables and OHLC without crude oil as exogenous variable respectively. The forecasting with crude oil as an exogenous variable on an average performs better than forecasting without crude oil as exogenous variable.

Model	Company	RMSE	Co-integration	Profit
LSTM	IOC	3.867	0.0105	39.556
	ONGC	2.804	0.0225	19.2896
	TCI	7.722	0.0006	41.5398
ARIMAX	IOC	1.3192	5.31E-11	676.0001
	ONGC	1.3109	6.13E-13	523.0499
	TCI	3.4484	1.61E-29	2331.6511
SVR	IOC	3.5724	5.01E-05	460.099
	ONGC	2.0611	0.0013	451.199
	TCI	7.3885	0.1454	1617.300

TABLE I
MODEL EVALUATION WHEN USING OHLC + CRUDE OIL

Model	Company	RMSE	Cointegration	Profit
LSTM	IOC	3.755	0.0068	51.3874
	ONGC	2.835	0.0102	23.6565
	TCI	8.247	0.0015	41.3388
ARIMAX	IOC	1.2971	5.57E-11	668.4251
	ONGC	1.3470	3.94E-18	523.0499
	TCI	9.6327	1.34E-29	1792.6009
SVR	IOC	5.9940	0.2405	3312.150
	ONGC	4.2434	4.69E-26	447.200
	TCI	13.717	1.37E-29	845.201

TABLE II
MODEL EVALUATION WHEN USING OHLC ONLY

B. Trend Analysis

As discussed earlier, correlations are found by calculating the percentage of the time series that follows a hypothesis. Table III gives the correlation score and a corresponding hypothesis. Here a tuple of the form (x,y,z) means that during a no trend period , the industry trend was hypothesized as x, during an up trend period, the industry trend was hypothesized as y and during a down trend period, the industry trend was hypothesized as z. The numbers 0, 1, -1 refer to ”no” trend, ”up” trend and ”down” trend respectively.

Industry	# Companies	Time Period	Hypothesis	Score
Agrochemicals	19	Yearly	(0,0,1)	0.7143
Plastic Products	38	Yearly	(0, -1, 1)	0.7143
Shipping	10	Yearly	(0, -1, 1)	0.7143
Cement	34	Half-Yearly	(0, 1, -1)	0.7143
Logistics	8	Yearly	(0, 0, 1)	0.8571

TABLE III
TREND ANALYSIS ACROSS INDUSTRIES

V. Conclusions

We trained multiple models to forecast the closing price of stocks of various companies that are affected by the changing prices of crude oil and also performed an analysis of the trends.

A. Forecasting

The observations made from the results are:

1) Consideration of crude oil in the analysis actually improves the forecasting.
2) ARIMAX performs better than any of the other models, which may be due to the following reasons:(i) The stock prices were taken on a day-to-day basis, and not by a minute-by-minute basis. The stock prices taken on a minute-by-minute basis usually brings in non-linearity in the time series due to minute fluctuations. A day-to-day change abstracts these minute fluctuations, thus making the series more linear. Since ARIMAX is a linear model, it forecasts well in this use-case. (ii) Deep Neural Networks are known to work for large amounts of data, but due to the previously mentioned constraint of not having enough data-points for the analysis explains why LSTM does not perform better than ARIMAX. (iii) The RMSE of Stock data using Open,High,Low prices using SVR is comparable to other model's errors but on using Crude oil attribute, it was found that SVR model does not perform well. Running Algorithmic Trading on the data obtained using SVR gave high deviation from the actual profit.

B. Trend Analysis

1) Agrochemical Industry - This industry thrives when there is a down trend in crude oil commodity prices. Probably because agrochemicals like pesticides which use petroleum byproducts as their key ingredient reduce their production cost as crude oil price falls.
2) Plastic Products Industry - The cost of production increases as crude oil prices go up and results in the down trend exhibited by the industry and vice-versa.
3) Shipping Industry - The cost of operation would increase with increase in fuel price and hence the down trend. Similarly the fuel expenses would reduce with reduction in fuel prices and hence up trend.
4) Cement Industry - Cement uses coke as a raw material for its production. An increase in raw material cost would increase the selling price of cement and its products to compensate for the same. Hence up trend in crude oil shows up trend in this industry and vice versa.

5) Logistics Industry - This industry hugely benefits from the reduction in fuel prices and hence exhibits an up trend when crude oil price goes down.

From these insights we conclude that factoring in the changing prices of crude oil in forecasting stock prices provides valuable insight for companies regardless of whether they are upstream or downstream.

REFERENCES

[1] Abdulrahman Adnan Alqattan and Ahmed Alhayky. Impact of oil prices on stock markets: evidence from gulf cooperation council (gcc) financial markets. 2016.

[2] G Box and M Jenkins. Time series forecasting analysis and control. *San Franscisco: Holden Day*, 1970.

[3] Shuang Gao and Yalin Lei. A new approach for crude oil price prediction based on stream learning. *Geoscience Frontiers*, 8(1):183–187, 2017.

[4] Sajal Ghosh. Import demand of crude oil and economic growth: Evidence from india. *Elsevier*, 2009.

[5] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.

[6] Linda Kaufman Alex Smola Vladimir Vapoik Harris Drucker, Chris J.C. Burges. Support vector regression machines. 1997.

[7] Virendra Rai and Palash Bairagi. Impact of changes in oil price on indian stock market. 04 2014.

[8] S Sathyanarayana, SN Harish, and Sudhindra Gargesha. Volatility in crude oil prices and its impact on indian stock market evidence from bse sensex. *SDMIMD Journal of Management*, 9(2):65–76, 2018.

[9] KDM Soundarapandiyan and M Ganesh. An analytical view of crude oil prices and its impact on indian economy. *IOSR Journal of Business and Management (IOSR-JBM)*, pages 23–28, 2017.

[10] Peter Whitle. *Hypothesis testing in time series analysis*, volume 4. Almqvist & Wiksells, 1951.

[11] Wen Xie, Lean Yu, Shanying Xu, and Shouyang Wang. A new method for crude oil price forecasting based on support vector machines. In *International Conference on Computational Science*, pages 444–451. Springer, 2006.