International Journal of Business and Management Sciences **E ISSN**: 2708 – 4337 **P ISSN:** 2708 – 4329
Available online at http://www.ijbms.org
International Journal of Business and Management Sciences
Volume 04 (03), 2023
Received, 12 July, 2023, *Revised 10 August, 2023,*
Accepted, 13 August, 2023, *Online First, 17, August, 2023.*

FORECASTING STOCK RETURNS: APPLICATION OF MACHINE LEARNING ALGORITHM TO PSX

[1]*Unbreen Arif,* [2]*Naveed Khan,* [3]*Junaid Rauf Burki*

Keywords:
Stock returns, forecasting, investment decisions, Financial Markets, Machine Learning.
JEL Classification:
G17, C53, G11, E44, M15

ABSTRACT

The Stock Exchange performance is reflector of financial health of economy, while prediction of stock returns considered to be a complex task. The stock returns are determined by many factors from company specific to the macroeconomic indicators and also dependent on behavioral factors associated with the investors sentiments. The emergence of Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized the traditional statistical approaches to forecast returns. The Current study is an attempt to forecast stock returns using ML Algorithm "prophet" to analyze performance of the prediction model via comparison of forecasted returns with actual returns. The model is implemented and tested in emerging market stock returns where the returns are highly volatile. For the purpose of analysis, data of all the firms listed in Oil and Gas sector in PSX were selected w.e.f. 2012 to 2021.The data distributed in training and testing samples to forecast returns with prophet model using python. The model performance is evaluated with evaluation matrix of MAE, MSE and RMSE. The results of the study indicated that the OGDC stock has reported superior performance of ML algorithm to forecast returns with MAE 0.002, MSE 0.0001 and RMSE 0.0108. The R^2 98% indicates that the Machine learning prophet model has greater ability to predict returns as the algorithm provides flexibility to capture trend, seasonality and holidays effect to forecast results according to the analyst requirements. The findings of study are useful for the fund Managers, investors and researchers to analyze trend and make optimal investment decisions.

INTRODUCTION

Stock price prediction is a complex task, traditionally statistics and econometrics models are used for forecasting of stock prices and to identify the pattern of returns with time. The selection and development of best model is extremely essential in order to forecast the stock

[1] *Assistant Professor, University of Education, Pakistan. unbreen.arif@ue.edu.pk*
[2] *Lecturer, Abbottabad University of Science and Technology, Pakistan. naveedkhan1988_2007@yahoo.com (Corresponding Author)*
[3] *Deputy Chief, MOL Pakstan Oil and Gas Co.B.V, Pakistan. Junaidburki@gmail.com*

price and to take appropriate decision. The opening, closing, days high-Low prices are significant variables that have correlation with the next days stock price and provide greater information to forecast next day price. The stock prices are dynamic and effected due to change in the risk factors attributed to interest rates, foreign exchange rates, prices of real-estates, equity price etc. (Boudabsa & Damir Filipovic', 2021). Moreover, the prediction of stock prices with the traditional models is challenging due to the limitation of stationarity as in the real world time series data observed properties of non-stationary. Moreover, ease to handle and to get reliable predictions from such time series data always required extensive analysis. The selection and development of best model is extremely essential in order to forecast the stock price in order to take appropriate decision. The opening, closing, days high-Low prices are significant variables that have correlation with the next days stock price and provide greater information to forecast next day price (Varian, 2014).

The stock prices are dynamic and to predict time series is a complex tasks. Many studies so far have used econometric models like Auto Regressive Integrated Moving Average (ARIMA) and Auto Regressive Conditional Heteroskedastic statistical models to forecast trends (Pascual, Romo, & Ruiz; 2006, Guo; 2019, Dinku et.al; 2022, Dong et.al; 2020). Since 1970 with the evolution of technology the use of IT tools and software are widely used to predict the stock trend in order to minimize risk and maximum return. Now a days several statistical models, machine learning and deep learning algorithms are used to predict trends.

The use of Machine learning is widely used in computer science and engineering fields for execution of complex forecasting tasks and to build causal analysis. Several studies have reported that with the application of Machine Learning the flexibility and the expectation of forecasted results from the complex data enhanced (Malladi, 2022). The availability of wide range of free software and ease of use has revolutionized the Machine Learning acceptance and make it a popular choice in field of finance as well. It is reported that ML algorithm provide more flexible approach to predict variable relationship in comparison to linear models for prediction of prices (Varian, 2014).

The current study is an attempt to implement the machine learning algorithm Prophet to predict the stock prices of Oil & Gas Sector firms listed in PSX. The Machine Learning algorithm is deployed on the real-world data of Pakistan Stock Exchange to identify trend and to predict the stock prices. To empirically evaluate the prophet model and for implementation of Machine Learning algorithm the closing prices of the selected stocks has been used as the target variable. The model is trained for making daily predictions of PSX data. The

performance matrix of the model is evaluated with MAE, MSE and RMSE scores. The results have led to conclusion that the prophet model is capable of predicting stock prices with reasonable accuracy. The main contribution of this paper is that it has empirically tested the forecasting ability of prophet model through utilizing the real time series data. The study will add in literature of use of Machine Learning to forecast stock prices trends and the results are useful for the investment Managers, fund managers and the researchers.

LITERATURE REVIEW

In financial market analysis forecasting of stock prices is the most complex and challenging due to multiple factors volatility impact. The changes in stock prices are attributed to sentiment of market, economic and political stability, investors behavior, trends in international market, and other macroeconomic factors. The volatility of returns, presence of risk and return tradeoff & time series data is required to use suitable model which provides reliable forecasted results. The accurate and reliable results required by the investors, fund managers and other stakeholders. The presence of multiple volatile factors required to effectively assess factors in a way to minimize forecasted errors instead of just relying on traditional statistical models of mean & variance.

Machine learning is defined as set of tools and techniques calibrated with computer science and statistics models to identify trends. It has been widely used and reported its success in engineering and computer science field. In 1974 Lee et.al (1974) introduced the machine algorithm in the field of economics, while in 1984 the Wang et.al (2019) has first time applied the machine learning algorithm for the research problem. There exist no uniform definition of Machine Learning it can be defined as the collection of multiple methods to analyzed big data and to forecast trends (Taddy, 2019). It is built around approximation techniques using to identify the predictive relationship through analysis of data. ML is successfully implemented in various task like fraud detection in banking industry and for financial planning and allocation of resources.

It is reported that the use of ML is extensively used for the recognition of patterns and trends (Wasserbacher & Martin, 2022). The trend in data is due to the existence of correlation among features which are used for the forecasting of results through utilization of big data. The sample size is very much important for the application of ML algorithm, small sample size lead towards estimation bias (Calainho et al., 2022) as for the calibration ML algorithm depends more on data. It is reported that traditional statistical do not work in presence of high dimensionality of data i.e in case of presence of large set of features are used to predict output

(Taddy, 2019). In 2018 (Gu, Kelly, & Xiu, 2018) has reviewed the research repository to justify the use of machine learning for asset pricing. The comparative analysis of various ML algorithms i.e. Dimension Reduction techniques, generalized linear model random forest and boosted regression techniques were revied and concluded the supremacy of Machine Learning Models for Asset Pricing over traditional Statistical techniques.

To differentiate role of Machine Learning to forecast trend and to build causal relationship the study conducted by (Wasserbacher & Martin, 2022). The review articled highlighted the use of Machine learning for financial planning and forecasting through use of Machine learning in Planning & forecasting. The results of the study differentiated the term forecasting and causal relationship and suggested to avoid the conflict in terms as the naïve use of just machine learning algorithms can be misleading and may provide unreliable forecasted results. Valuation of financial assets and to build optimal portfolios keeping in view of risk and return is always been a complex task. The risk and return determination and hedging is an integral part of financial and insurance business (Boudabsa et.al, 2021). To handle large sets of data and make valuation of Assets to decide whether to opt for the opportunity or not is dependent on the model ability to accurately predict results. The study result of (Calainho et al., 2022) shown that the ML yield higher accuracy of results to determine real state index returns. In the earlier Machine learning algorithms are suitable to predict or forecast time series of large data sets but now many new machine learning algorithm do not require extensive and large datasets to make future predictions (Gogas & Papadimitriou, 2021).

Weigand, 2019 highlighted the benefits of Machine learning in asset pricing through providing theoretical overview of latest studies deployed ML for various asset pricing like Equity, Bond, Derivatives and real estate and reported that the use of Machine Learning offer benefit through utilizing it in specific settings according to the requirement. For the prediction of stock prices through Machine Learning algorithm study on Dhaka Stock Exchange was conducted by (Islam et al., 2021). Through comparison of root mean squared error (RMSE) of ML algorithm i.e. Support Vector Regression and KNN(K-nearest neighbor) regression the study result reported that model performance of SVR was superior to KNN.

In order to assess the investors behavior on financial and social media news through Machine learning algorithm like random forest classifier, deep learning on the impact of stock prices. The 80.53% and 75.16 % stock price prediction accuracy was reported by (Khan et al., 2022). To forecast demand in phenomena of supply chain Management hybrid model Prophet and

SVR was tested by (Guo et al., 2021) and reported accuracy of the model to capture seasonality of time series data.

Pakistan is an emerging economy and its stock market is highly volatile. To predict stock prices is of high interest to the investors and fund managers so the deployment of Machine Learning algorithm in prediction of stock prices is required to be explored to analyze pattern and trend forecast of complex time series data. This study is an attempt to deploy ML algorithm prophet to forecast stock prices.

METHODOLOGY

The stock of Oil & Gas companies listed in PSX have been selected for the time period 2011 to 2021. The oil & gas sector stocks has been actively traded in the market and the great choice of investor so the first step is to fetch daily stock price data from the Pakistan Stock Exchange (PSX, 2022) and Yahoo finance (Finance, 2022) through appropriate choice of parameters. After getting data next step to clean it for the training objective. After splitting the data set into train & test Machine learning Algorithm Prophet is applied to time series data to predict the trend and forecast stock prices.

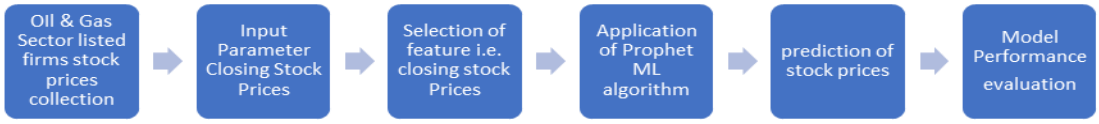

Fig- 1 The Flow Chart of Machine Learning Algorithm Prophet
Source: Author's own work

Prophet is an open source library from face-book for forecasting time series data in shortest span of time. It is process to predict the trends arrived from daily, weekly, seasonal & holiday effects of non-linear time series data. It has been in used to provide reliable forecast in many areas across facebook and has an ability to robust outcome where time series data depicts dramatic changes , missing value and have outliers (Robson, 2019) . The prophet has claimed accuracy, speed and the reliability of forecast as compare to other ML algorithms. It served great to the data where there is missing values, dirty data, outliers and shifts in trends (Kaninde et al., 2022) and ability to predict price in few seconds. Instead of just observing the time dependence trend the Prophet is actually "*framing the forecasting problem as curve fitting exercise*".

The mathematical model of prophet

The Prophet Model is an open source library from facebook team to forecast the time series data ((Taylor & Letham, 2018) . It has reported efficiency to predict results in shortest span of time. The prophet mathematical equation for time series data forecasting is composed of 3 components i.e. trend, seasonality and holidays. The equation can be written as under :-

$Y(t) = g(t) + s(t) + h(t) + e(t)$……. (1)

Where

g(t) = growth over time / trend (non-periodic changes)

s(t) = seasonality (daily / weekly / monthly) i.e. periodic changes

h(t) = holidays (irregular schedule)

e(t) = idiosyncratic changes

the g(t) in the prophet model can be take form of nonlinear saturating growth, linear trend with change points i.e. “piece-wise linear trend model”, atomic change point selection and forecast uncertainty. The selection of the trend term is dependent upon the characteristics of the dependent objects.

The linear trend in prophet model is expressed as

$g(t) = (k + a(t)^T \delta)\ t + (m+$

$a(t)^T \gamma)$……..……..(2)

for non-linear growth the model can be expressed as

$g(t) =$

$$\frac{C\ (t)}{1+\exp\ (-(k+a(t^T)\delta)(t-(m+a(t^T)\gamma))]}$$……..(3)

where

k = growth rate

δ = rate adjustment of growth rate

γ = vector of correction adjustment at change point

a(t) = vector of adjustment parameter

C(t) = time varying capacity

m = offset parameters

In time series seasonality is a pattern that repeat over the specific interval of time i.e. daily / weekly / monthly / yearly. The prophet model used Fourier Series to determine the periodic seasonality of time series. The mathematical equation is as under:-

$s(t) = \sum_{n=1}^{N}(a_n \cos(\frac{2\pi nt}{p}) + b_n sin\ (\frac{2\pi nt}{p})$)……(4)

for fitting the seasonality the estimation of 2N vector is required and it is assumed that in Prophet Model that

$\beta = [a1, b1, \ldots \ldots \ldots aN, bN]^t$……..……..(5)

Moreover $\beta \sim Normal\ (0, \sigma^2)$ prior to seasonality to impose smoothing of data.

The h(t) are holiday events in the prophet models they are predictable events, unlike to above mentioned factors they are predictable events and have similar effects on the data every year. They are required to incorporate in the time series data for the accurate forecasting. It is assumption of the model that each holiday have an independent effect on the forecasted results. The holiday effect can be expressed as under :

h(t) =

Z(t)K…….(6)

where Z(t) = [1(t ϵ D_1),……1(t ϵ D_L)]

& K$\sim Normal\ (0, v^2)$

Where

D_i = past & future dates of holidays

K = parameter to capture forecasting change in the time series due to holidays

In this study the Prophet Model is implemented to Oil & Gas Sector Companies stocks listed in PSX. Pakistan is an emerging economy and reliant on the energy sources of oil, gas & hydel energy. The fluctuation in the international oil market prices have direct impact on the economy of Pakistan and also reported significant impact on the profit of the firms and Pakistan stock market. The stocks of the Oil & gas Exploration and production companies are the first choice of the local and international investors moreover a drop in price of oil have a multi- million impact on economy of Pakistan (Hussain, 2016). These sectors are also exposed to the seasonality the uncertainty and instability of stock prices due to seasonality is a complex and great challenge to the investors and the fund managers and have an impact on the buying and selling decision of the stocks. To determine the forecasted prices, the prophet model is deployed on the stock prices of MARI, OGDC, POL & PPL from January, 2011 to December 2021 with total of 2232 observations.

RESULTS

The prophet is implemented in python by using prophet library. The daily frequency is chosen to capture the trend of time series and data is determine the trend identify the best model to determine the forecasted price of stocks of this sector is of great interest to the fund

managers, regulators and local and international investors. This study is an attempt to deploy the return associated with the trading stock prices. So, the interest of the investors is strongly associated with the prices of the stocks.

The data is trained by using fit function available in prophet. First, named the target variable i.e. close price as y and features i.e. Dates as ds. The EDA is performed to see the return and variability of time series data.

Table-I EDA of Oil & Gas Stocks

Statistics	MARI	OGDC	POL	PPL
Count	2232	2232	2232	2232
Mean	820.54	167.78	443.69	170.26
St. Dev	541.93	47.55	99.104	40.23
Min	81.45	75.01	189.67	69.13
25%	246.99	137.79	374.00	147.54
50%	876.09	155.69	432.78	175.30
75%	1344.16	189.75	507.34	201.29
Max	1809.41	287.84	707.34	260.06

Source: Research findings

The figure 2 indicate the historical trend of the stocks prices of the oil and gas sectors stocks. It is depicted from the graph the returns are volatile over the study period from 2012 to 2021.

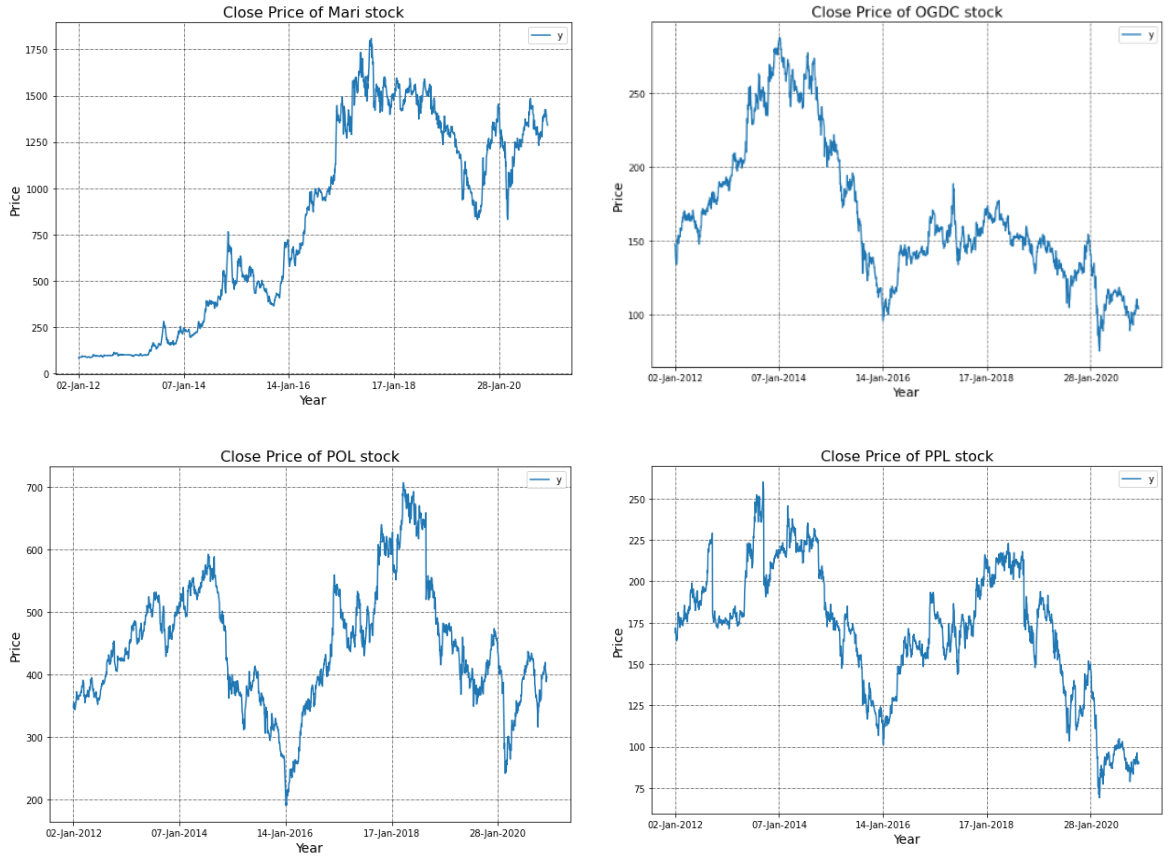

Fig-2 The stock Prices Trend from 2012 to 2021
Source: Research findings

The graph is plotted by taking dates on the x-axis and the stock price data of all the firms stock listed under Oil & Gas sector in Pakistan Stock Exchange. The data range from January 2012 to December-2021.

The prophet model is fit for the prediction of the stock prices for the next one year. The original prices along with one year forecasted prices of all selected stocks is presented via graph as under:-

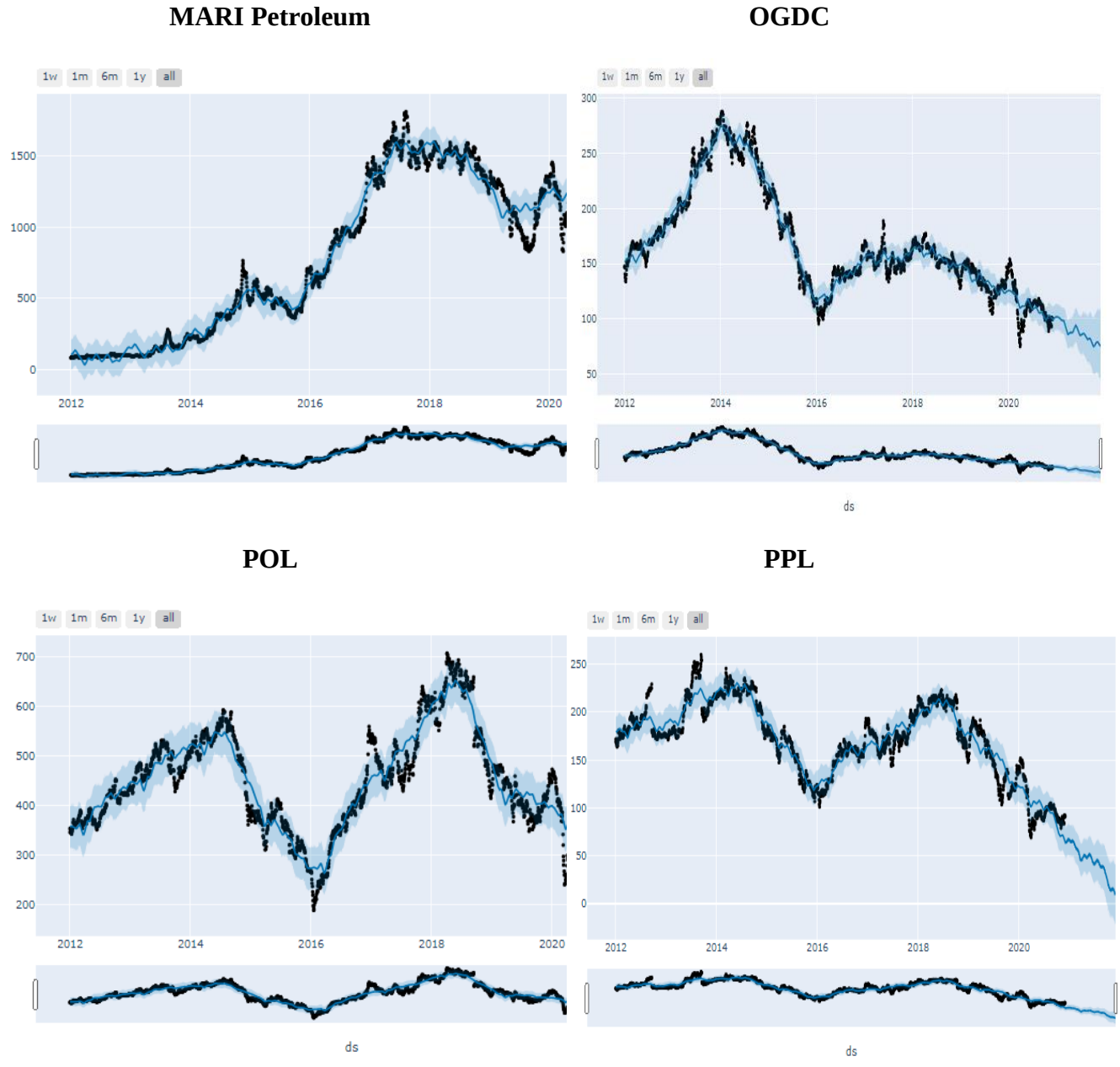

Figure-3 Predicted Trend of Stock Prices Oil & Gas Sector Stocks
Source: Research findings

In order to check the prediction accuracy the predicted values are then compared with the actual prices of that day. The sample predicted price alongwith their actual price on that day

is presented in below table. The forecasted values are approximately equal to the actual observations and have shown the predicted accuracy of the model as under:

Table-II Comparison of forecasted Price Vs Actual Prices of Oil & Gas Sector stocks:

Ticker	Date	Actual Price	Forecasted Price
MARI	03.12.2020	1339.53	1390.02
OGDC	03.12.2020	99.59	100.1
POL	03.12.2020	348.54	401.61
PPL	03.12.2020	91.38	64.88

Source: Research findings

In order to obtain performance matrix of prophet Model and to measure the accuracy of forecast the values of Mean Absolute Error MAE, Mean Squared Error MSE, Root Mean Squared Error RMSE were calculated for all the selected stocks

Table-III MAE, MSE & RMSE of Oil & Gas Sector Stocks of PSX

Ticker	MAE	MSE	RMSE	R^2
MARI	0.0226	1.1421	1.0687	90.27%
OGDC	0.0002	0.0001	0.0108	99.99%
POL	0.0238	1.2681	1.1233	99.97%
PPL	0.0119	0.3146	0.5609	99.98%

Source: Research findings

The performance and accuracy of the Machine Learning Model prophet has analyzed with four indices Mean absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-squared. The results were obtained by using two value sets i.e. the actual observations and the predicted values of prophet model. The MSE measures the average of the absolute difference between the actual value magnitude of errors forecast. The results presented in Table-III indicated that the lowest value of the MAE 0.0002 is reported by OGDC stock while all other stocks have also reported value close to zero. The Mean Squared Error (MSE) is the sum of the squared deviation from predicted to actual observations and RMSE is square root of MSE the table-III results of MSE and RMSE also reported the lowest value for all the stocks of Oil and Gas Sector which is an indication of model accuracy the R-squared of above 90% indicates the predicting power of the prophet model.

DISCUSSION

The current study has utilized prophet methodology to forecast returns of oil and gas sector stocks. The forecasting of stock returns is a complex task and the results are dependent on the techniques used to forecast returns. The use of Machine learning has revloutioned the forecasting techniques and several prediction algorithms are used with reported efficiency to forecast stock returns. To assess the forecasts, statistical indicators are computed on the testing sets.The outcomes of the forecast assessment in Figure-3. The Each indicator value should be as low as possible.

the model's effectiveness. The results shown that the selection of stocks gets easier for the investor on the basis of the results derived from Facebook prophet in order to create optimal portfolio of investment. The prophet model provides flexibility to capture trend and to account for seasonality, holidays as per the requirement to analyst and per actual scenario to forecast results. The results are in confirmation of the studies of Garlapati, Krishna, & Narayanan (2021) and Huang (2022) of superiority of prophet model to forecast stock returns. The result of evaluation matrix and the visuals of figure-3 indicates that the proposed ML facebook model has predicting power to forecast returns in emerging market of PSX. The inclusion of other available feature of prophet model can enhance the accuracy of the results and make it convenient application to use for the prediction of prices. The current study is deployed on oil and gas sector of PSX the stocks of oil and gas sectors are actively traded stocks in PSX and due to fluctuation of oil prices and change in macroeconomic factors have also impacted on the returns of the stocks , The use of ML is efficient and widely used for the recognition of patterns and trends (Wasserbacher & Martin, 2022). So for active trading stocks the result of study can be useful for implementation of prophet model to forecast returns and optimal investment decisions.

Conclusion

The forecasting of returns is a complex and challenging task the development of AI and use of Machine Learning Algorithm has revolutionzed the traditional forecasting process. The current study has implemented the prophet model a machine learning algorithm to predict the emerging market returns. The results indicates that the model provides flexibility to capture trend and to account for seasonality, holidays as per the requirement to analyst and per actual scenario to forecast results.

Implications and Future directions

The stock price prediction is always considered a challenging task for the investors and fund managers the selection of stocks for the construction of portfolios is of significant importance to earn optimal returns. Fund manager used fundamental and technical analysis by using firm level and market data to analyze returns and risk. the use of Machine leaning with the evolution of Data Science has evolved the traditional forecasting techniques due to their reported efficiency and performance of forecasting returns. The current study results are useful for the fund managers decision makers and the investors to predict stocks returns to get optimal portfolio returns. The current study utilized data set of Oil & Gas sectors stock which

are actively traded in PSX on ten years daily stock returns. so, in future studies may also be initiated on large datasets for the validation of the prediction power of prophet through Machine Learning and for the evaluation of value investment stocks. Moreover, in future studies may also be initiated to compare the prophet model results with other Machine Learning algorithms both supervised and unsupervised.

REFERENCES

Boudabsa, L., & Damir Filipovi´ c. (2021). *Machine learning with kernels for portfolio valuation and risk management*. *26*, 131–172.

Calainho, F. D., Alex M. van de Minne, & Marc K. Francke. (2022). *A Machine Learning Approach to Price Indices: Applications in Commercial Real Estate*. 30. https://doi.org/10.1007/s11146-022-09893-1

Dinku, T., & Worku, G. (2022). Asymmetric GARCH models on price volatility of agricultural commodities. *SN Business & Economics, 181*. doi:https://doi.org/10.1007/s43546-022-00355-7

Dong, H., Guo, X., Reichgelt, H., & Hu, R. (2020). Predictive power of ARIMA models in forecasting equity returns: a sliding window method. *Journal of Asset Management, 21*, 549-566.

Finance, Y. (2022). *Yahoo Finance*. Retrieved from https://finance.yahoo.com/.

Garlapati, A., Krishna, D. R., & Narayanan, G. (2021). Stock Price Prediction Using Facebook Prophet and Arima Models. *6th International Conference for Convergence in Technology (I2CT)*. doi:DOI:10.1109/I2CT51068.2021.9418057

Gogas, P., & Papadimitriou, T. (2021). Machine Learning in Economics and Finance. *Computational Economics, 57*(1).

Gu, S., Kelly, B., & Xiu, D. (2018). Empirical asset pricing via machine learning. *Working Paper, University of Chicago*.

Guo, Z.-Y. (2019). EMPIRICAL PERFORMANCE OF GARCH MODELS WITH HEAVY-TAILED INNOVATIONS. *Bulletin of Economic Research, 71*(3), 359-387.

Huang, Q. (2022). Forecasting Stock Prices Using Multi-Macroeconomic Regressors Based on the Facebook Prophet Model. *BCP Business & Management, 25*. doi:DOI:10.54691/bcpbm.v25i.1762

Hussain, D. (2016, january 18). *Impact of Slippery Oil on PSX Sectors*. Retrieved from dawn.com: https://www.dawn.com/news/1233741

Islam, S., Sikder, Md. S., Hossain, Md. F., & Chakraborty, P. (2021). Predicting the daily closing price of selected shares on the Dhaka Stock Exchange using machine learning techniques. *SN Business & Economics, 1*(4), 58. https://doi.org/10.1007/s43546-021-00065-6

Khan, W., Mustansar Ali Ghazanfar, Azam, M. A., Amin Karami, Khaled H. Alyoubi, & Ahmed S. Alfakeeh. (2022). *Stock market prediction using machine learning classifiers and social media, news*. 3433–3456.

Lee, C., & Lee, E. (1974). Fuzzy sets and neural networks. *Journal of Cybernetics, 4*(2), 83-103.

Malladi, R. K. (2022). Application of Supervised Machine Learning Techniques to Forecast the COVID-19 U.S. Recession and Stock Market Crash. *Computational Economics*. doi:https://doi.org/10.1007/s10614-022-10333-8

Pascual, L., Romo, J., & Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis, 50*(9), 2293-2312.

PSX. (2022). *Data Portal PSX*. Retrieved from https://dps.psx.com.pk/: https://dps.psx.com.pk/

Riady, S. R. (2023). Stock Price Prediction using Prophet Facebook Algorithm for BBCA and TLKM. *International Journal of Advances in Data and Information Systems*. doi:DOI:10.25008/ijadis.v4i2.1258

Robson, W. (2019, June 17). *Future-Vision.* Retrieved from Medium.com: www.medium.com

Taddy, M. (2019). *Business data science: combining machine learning and economics to optimize, automate, and accelerate business decisions.* McGraw-Hill Education.

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician, 72*(1), 37-45.

Varian, H. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives , 28*(2), 3-28.

Wang, H., Li, C., Gu, B., & Min, W. (2019). Does AI-based credit scoring improve financial inclusion? Evidence from online payday lending. *40th international conference on information systems.* ICIS 2019.

Wasserbacher, H., & Martin, spindler. (2022). *Machine learning for financial forecasting, planning and analysis: Recent developments and pitfalls*. *4*, 63–88.

Weigand, A. (2019). *Machine learning in empirical asset pricing*. *33*, 93–104.