Article

Access Risk Management for Arabian IT Company for Investing Based on Prediction of Supervised Learning

Bhupinder Singh [1,*] and Santosh Kumar Henge [1]

[1] Department of Computer Science and Engineering, Lovely Professional University, Phagwara (144401), Punjab, India

* Correspondence: bhupinder0741@gmail.com

Abstract: The study focuses on chances of profit from Saudi IT company to increase with few losing trade and a less margin winning investing decisions. Fear and greed are two psychological points that dominates the investing decisions. The main objective of the research to study the risk management related to Al Moammar Information Systems that is listing on Saudi Share market. Previous Research relied on limited methods for prediction of accurate price for investing in the current bullish Markets. The research also emphasizes on predicting the right price for investing on the basis of Supervised Learning methods involving Support Vector Machine, Random Forest Regression, XGBoost, Auto Arima and Quasi Poisson Regression. Research has found that the right price to investing in this company comes out to be 106.945 on the prediction of previous 6 months period data. Data is sourced though Yahoo Finance api in form of Date, Open, High, Low, Close, Volume, Dividends and Stock Splits. This solution can be fruitful for newly trained investors who are willing to invest for long term basis.

Keywords: Support Vector Machine; Random Forest Regression; XGBoost; Auto Arima; Quasi Poisson Regression; Risk Management

1. Introduction

Stock market risk management has played a predominant role in surviving in the volatility scenario. Saudi stock Markets have revolutionized the flow of trend in the International Market. Supervised Learning is focused on predicting the ideal price for the investing in the bullish market. Al Moammar Information Systems has revolutionized the Saudi Stock Market by showing 100-percentage growth in consecutive years. The problem arises when new investing is struggling to find the appropriate stage to enter into the market. Predicting the stock based on sentiment analysis [5] leads to the complexity of the model and thus makes it difficult to take a decision in financial stock market. Risk management based on previous knowledge may increase in uncertainty that leads to financial loss in day trading or long-term investments. Previous Studies indicates that investors relied on the fundamental aspects of the companies but Machine based supervised Learning [9] is a booster for effective as per the technical data of the company. Data calibrate for one year for risk management and further technical prediction has used for latest six month of data to increase the execution time of the system. Previous year's data has formulated as the training data [6]. The yahoo finance symbol of Al Moammar Information Systems is 7200.SR and yfinance model is required to install as external

library in google collab framework. Classification of various supervised learning are represented in form of flowchart.
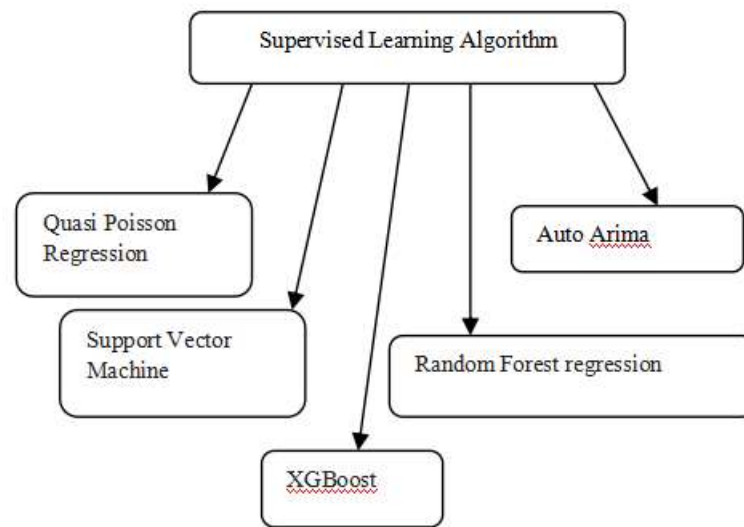

Fig. 1 Supervised Learning Methods.

A. Related Work

D. Wei (2019) accessed the role of time series in historical information. He also stated that LSTM model has some limitations such as time lag for prediction [2]. M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar (2020) considered the Artifical Neural based Network and Random Forest methods for next day closing price of different sectors of companies. They stated that low values of RMSE, MAPE concluded that models are quite efficient for prediction [5]. H. R. Patel, A. M. Patel and S. M. Parikh (2020) concerned about the role of Machine Learning to find the particular knowledge from uncertain data by analysis of different methods [6]. Seungho Baek, Sunil K. Mohanty and Mina Glambosky (2020) focused on understanding the change related to volatility using Markov Switching AR model. They also confirmed that COVID-19 news affected the stock prices of company [7]. A. F. Wagner (2020) elaborates the after effects of COVID-19 on the complexity of the company profit or loss ratios in relation to expected cash flows [9]. D. Shah, H. Isah and F. Zulkernine (2019) provided the review for the different taxonomy methods of stock prices movements and discussed various challenges associated with machine learning methods [10]. Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, Tat-Seng Chua (2019) deployed adversial training to enhance the generalization of the prediction model and proposed to add perturbation to implement price variable to work well [15]. E. Letizia and F. Lillo (2019) implemented predominant correlation among the local topology attributes of firm and their associated risks [17]. H. Shah (2019) concluded that Levenberg Marquardt methods perform better for prediction than the ANFIS in terms of memory allocation and accuracy [20].

2. Simulation of Risk Management

The investment in stock market requires assuming a certain amount of risk. Intelligent Investor [19] utilizes efficient methods to minimize the market risk and the boost the gains associated with the market trend. The biggest challenge is to find the appropriate stage to enter into the market in the bullish markets. There exists an enormous risk for an over scribed levels and possess higher chances

to come down to some extent. Table 1 shows the evaluation of average closing price of the stock for N number of days [7] as calibrated in Al Moammar Information Systems stock price. The average price has calculated for 1 week, 2 weeks, 1 month, 3 months, 6 months and 1 year eventually. Author has focused on following the trend of the direction of stock market and stop loss leads to financial loss.

Table 1. Evaluation of average closing price and profit or loss for the N number of days.

Period	Average Price	Profit or Loss
1 week	50.62358038766043	6.903356352695258
2 weeks	50.89297458103725	-2.497292811798067
1 month	54.29145151774089	9.923665858258898
3 months	63.86256421407064	6.534656256543506
6 months	84.87345790863037	4.550047454134997
1 year	102.67734532129197	49.826950441055025

Furthermore, the Table 1 calculate the profit or loss for the N number of days. The profit/ loss has calculated for 1 week, 2 weeks, 1 month, 3 months, 6 months and 1 year respectively.

Table 2. Evaluation of average closing price and profit or loss for the N number of days.

Daily Returns	Trend
- 0.5 && 0.5	very slight change or no change
0.5 && 1.0	slight change on the positive side
-3 && -1	slight change on the negative side
1 && 3	change on the positive side
-3 && -1	change on the negative side
3 && 7	top gains
-3 && -7	top losses
> 7	**bull run** (stock prices are on rise)
< -7	**bear** (stock prices are on decline)

Next, the trends of the stock are perceiving based on the daily returns i.e., daily percentage change in closing price of particular stock. The trends analyzed based on the following factors such as dependable parameters and non-dependable parameters.

A. Outline of Bear and Bull Market

The words bull and bear market used to explain how financial prices usually do whether or not they increase or depreciate. At the same time, given that the economy is driving by the behavior of consumers, these words often reflect how investors feel about the business and the patterns that follow.

Table 3. Evaluation of trends of the stock is observed based on the daily returns.

Trend	Mean	Median	Counts
Among top gainers	1.039402e+06	835992	27
Among top losers	7.932692e+05	587088	11
Bear	1.463943e+06	1463943	2
Bull run	1.747939e+06	1585293	7
Negative	2.770925e+05	226514	42
Positive	4.743291e+05	376900	42
Slight Negative	2.522934e+05	171047	27
Slight or No Change	2.189553e+05	144980	62
Slight positive	2.998117e+05	219735	31
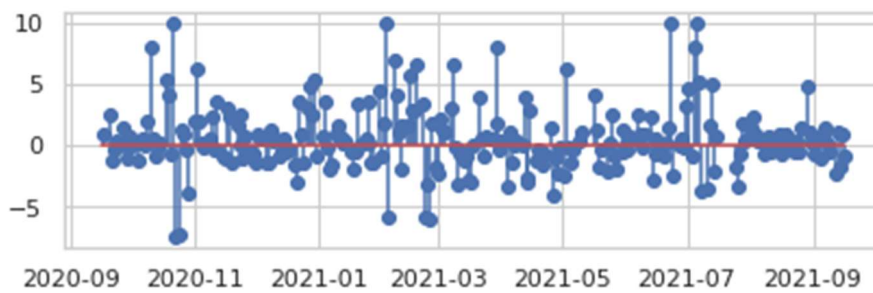
Fig. 2 Exhibit stem plot in form of discrete series.

The above graph is a stem plot, which is a discrete series plot. It helps to gain insights from the dataset and highlighting the outliers in the dataset [22]. The daily returns of the closing price of the stock are plotting over the time. If the 14-day rolling average is closer to upper band, it indicates that the stock overbought whereas if it closer to lower band it indicates that the stock is oversold [12]. The above plots are bar graphs. It plots the mean and median respectively of the total traded quantity of the stock grouped by the trend it follows.
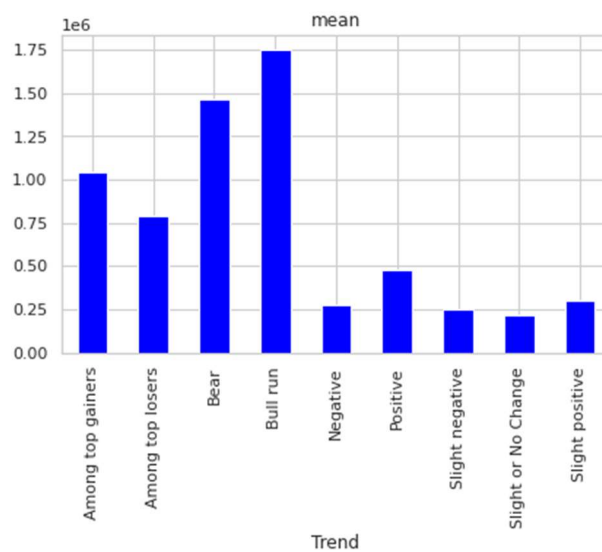

Fig. 3 Shows the mean of the total traded quantity of the stock grouped by the followed trend.
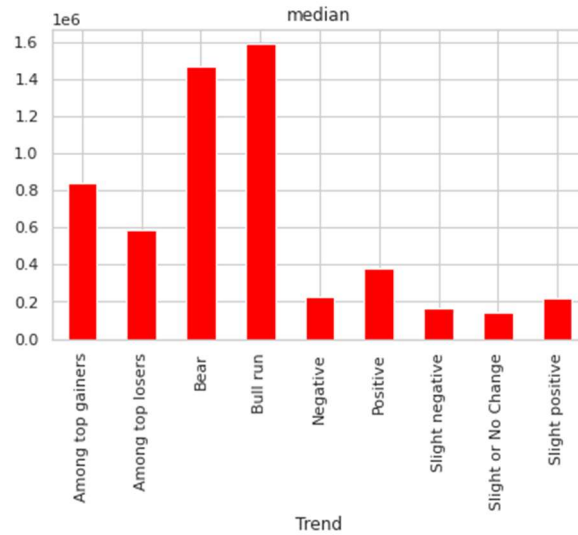
Fig. 4 Shows the median of the total traded quantity of the stock grouped by the followed trend.

The Fig.5 shows the percentage of each type of trend in the stocks. The trend constituting the highest percentage is the slight or no change in the daily returns of the stock indicating that the stock is non-volatile [15].
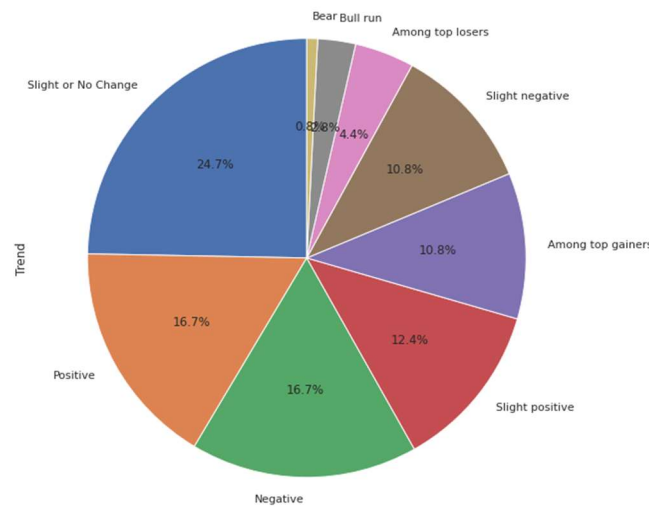

Fig. 5 Evaluation of average closing price and profit or loss for the N number of days.

$$Upper\ Band = MA + Constant \sqrt{\frac{\sum_{i=1}^{n}(y_i - MA)^2}{n}} \tag{1}$$

$$Lower\ Band = MA - Constant \sqrt{\frac{\sum_{i=1}^{n}(y_i - MA)^2}{n}} \tag{2}$$

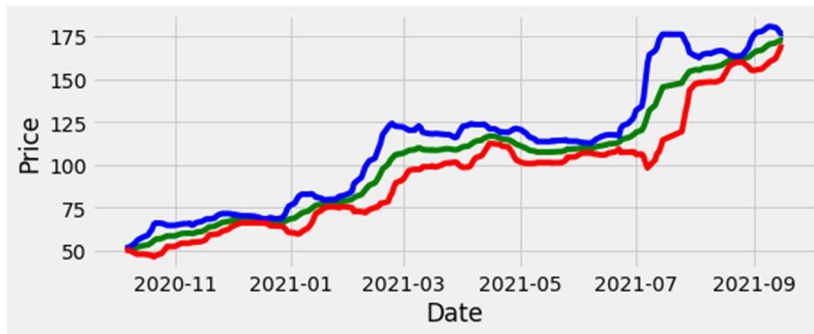where MA is Moving Average.

Fig. 6 Shows plot of Bollinger Bands.

The above plot is of Bollinger Bands. Bollinger Bands are a type of statistical chart characterizing the prices and volatility over time period for given interval. Equation 1-represented computation of Upper band and equation 2 is labeled as Lower band while MA is moving Average it consists of 3 bands described: 14 day rolling/moving average as green band; Upper Band (blue band) **-** Calculated using the addition of 2 times the standard deviation of the 14-day rolling Standard deviation from the 14-day rolling mean.
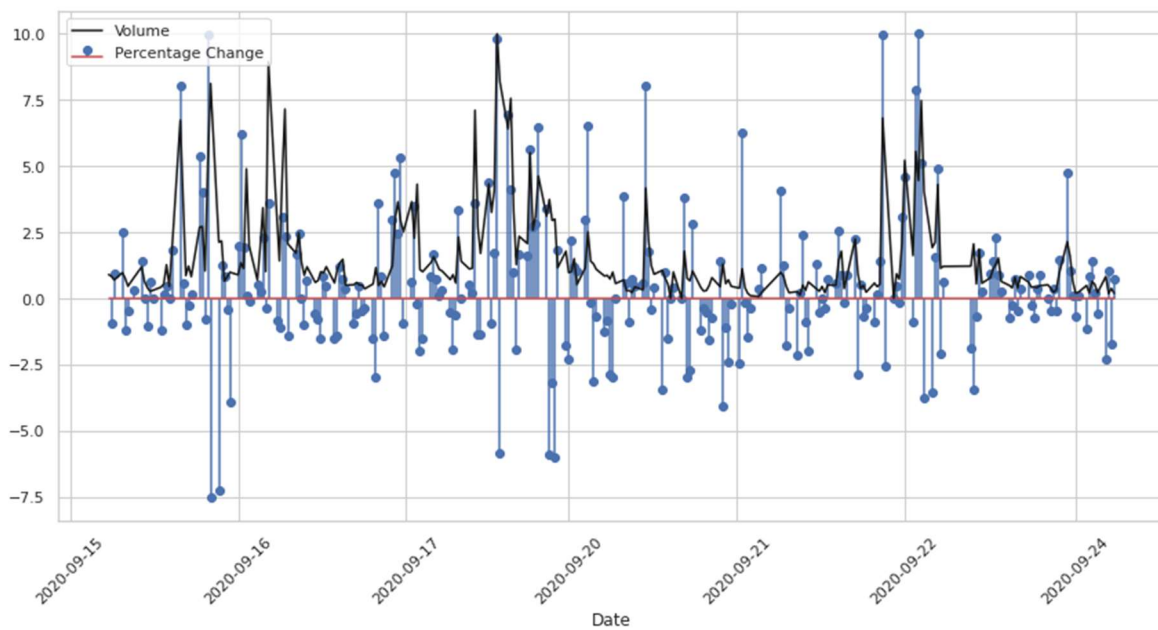

Fig. 7 Describes the total traded quantity of the stock for the entire timeframe over the stem plot of daily returns.

Lower Band (red band) **-** Calculated using the subtraction of 2 times the standard deviation of the 14-day rolling standard deviation from the 14-day rolling mean. The Fig.7 plots the total traded quantity of the stock for the entire timeframe over the stem plot of daily returns. It helps to derive a relation between the daily returns of the stocks and the total traded quantity. As it is visible in the plot, the total traded quantity usually increases when the daily returns are high indicating a positive relationship [13] but it not always trues in some parts. Table 4 observed the structure of category of parameters with respect to its role in volatility of share price and dependency of its value alter the situation of investor mindset to enter into the market [8].

Apart from the stock value variations, there are certain parameters are associated with stock market environment.

Table 4. Describes the structuring of dependable parameter and non-dependable parameters.

| Dependable Parameters	Non-Dependable Parameters
Invested Budget (IB)	Political crises (PC)
Number of Projects In hand (NPH)	Climate and Uncertainty (C-UC)
Number of Projects take over in future (NPTF)	Data Leakage (DL)
Manpower (MP)	War inside Country (WiC)
Infrastructure (In)	Brand image Worldwide (BIW)
Oil Prices (OP)	Company CEO instability (C-CEO-I)
Electronic Prices (EP)	Corporate Governance (CG)

3. Evolution of Supervised Learning
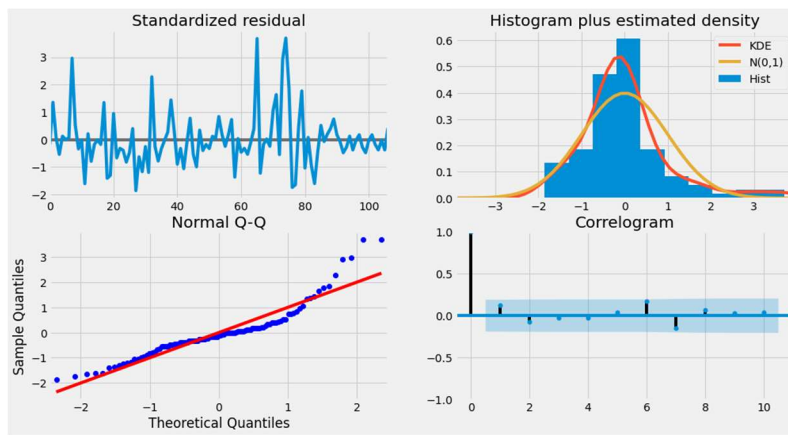
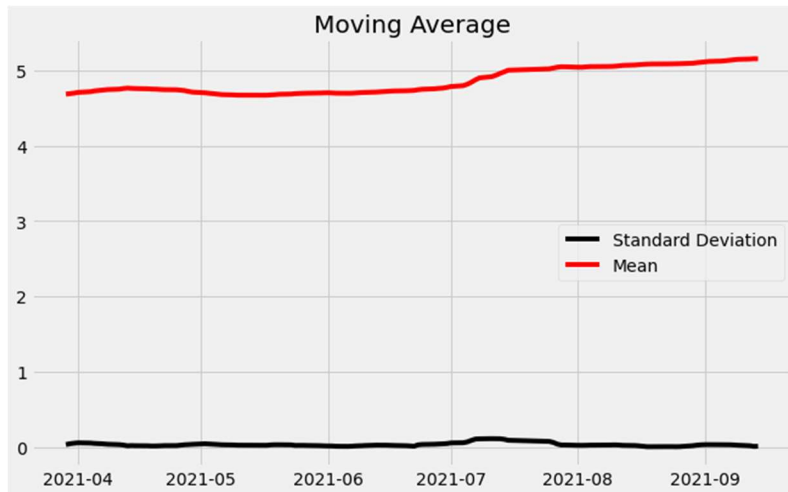**Fig. 8** Shows time series forecasting in form of correlogram.

Fig. 9 Describes the modest variation of mean and standard deviation.

The practical approach to machine learning is focusing using supervised learning. Commonly, Input variables (x) and output variables (y) are implementing into mapping function $f(x)$ to predict the future predictions. Learning concludes when appropriate method achieves desired level of performance. Initially, time series forecasting is done to study the pattern of data into four

decompositions such as Trend, Seasonal Variations, Cyclic variations and random movements. Fig.8 is representation of the time series forecasting of share price.

$$MA_k = \frac{p_{n-k+1} + p_{n-k+2} + \ldots + p_n}{k} = \frac{1}{k}\sum_{i=n-k+1}^{n} p_i \tag{3}$$

where $p_1, p_2, \ldots, p_n$ are the data points that are considered as closing price.

$$s = \sqrt{\frac{1}{N-1}\sum_{I=1}^{N}(x_i - \bar{x})^2} \tag{4}$$

Where N-1 represented degree of freedom, $(x_1 - \bar{x}, \ldots, x_n - \bar{x})$ are the deviations from the mean and Fig.9 shows the modest range of comparison between standard deviation with respect to Mean.

Support vector machine can be considered best-supervised learning method for both classification and regression problems. Normally, support vector machine [9] builds hyper plane in a highly space in form of dimensions. New data point has to decide which class has to be included called linear classifier. Author has implemented Linear SVM.

Let assume a training data set of n points in $(x_1, y_1), \ldots, (x_n, y_n)$ where y_i are classified as 0 or -1 which belongs to each point x_i. Thus, Hyperplane can be expressed as for given set of points x.

$$W^T x - b = 0 \tag{5}$$

Where w is the normal vector with respect to the hyperplane. Further move, Compute the Support Vector Machine Classifier minimizing a normal form. Though, Precision in equation 7 anticipates about how much model is accurate in form of predicted positive. Precision is an excellent way to signify the cost of high value of false positive. Recall in equation 8 evaluates how many of Actual Positives our method simulating through labeling it as bests case True Positive [14].

$$Classifier = [\frac{1}{n}\sum_{i=1}^{n} max\,(0, 1 - y_i\,(W^T x_i - b))] + \gamma||w||^2 \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

Where *TP* is true positive, *FP* is false positive, *FN* is false negative. The normal F1-score measure as 0.483415 and jaccard similarity score is 0.48. Author has realized average F1-score is 0.3933 and Jaccard score is 0.4400. F1-score as label in equation 9 is required to study the exact balance between recall and precision [10].

Table 5. Evaluation of accuracy of SVM in terms of Precision, recall andf1-score.

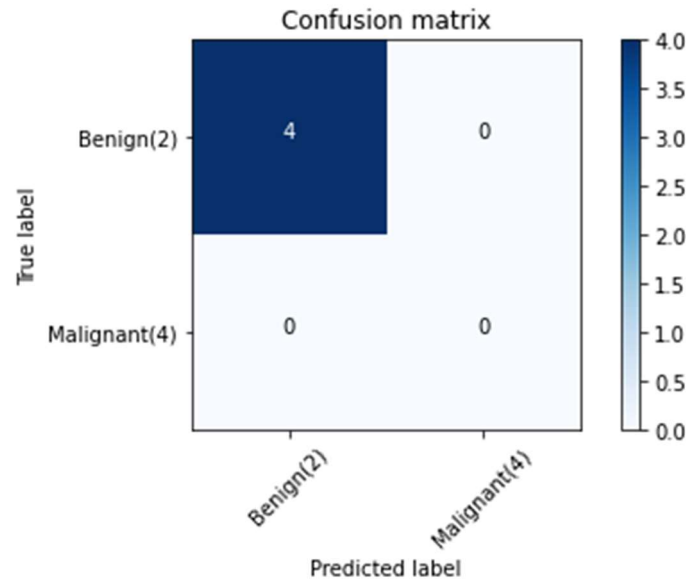|	Precision	Recall	F1-score	Support
0	0.57	0.53	0.55	15
1	0.36	0.40	0.38	10
accuracy			0.48	25
macro avg	0.47	0.47	0.47	25
weighted avg	0.49	0.48	0.48	25

Fig. 10 Shows the confusion matrix for support vector simulation with respect to predicted label.

Random Forest regression is most flexible machine learning algorithm that does not require hyper tuning is referred as Random Forest regression. One of the tiresome problems is over fitting of classifier data that is overcoming by this type of supervised learning. The Out of bag R-2 score estimate comes out to be 0.9948780115436955 and Test data R-2 score has calibrated as 0.9969415555766163.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{10}$$

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}} \tag{11}$$

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(\acute{x}) \tag{12}$$

After training, predictions [23] for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x': Random Forests Regression Score: 0.9969415555766165. Test data Spearman correlation is 0.9921138864039075 as computed in equation 10. Test data Pearson correlation is 0.9989167770310116 as evaluated in equation 12.

Xboost for time series forecasting [12] is used as automatic parametric selection of features and there exist a availability of extra parameters related to randomization. The version used is 0.90 in this experiment. Initialize model with a constant value as represented in equation 13.

$$\hat{f}_{(0)}(x) = \underset{\theta}{\mathrm{argmin}} \sum_{i=1}^{N} L(y_i, \theta) \tag{13}$$

$$Output\ f(x) = \hat{f}_{(M)}(x) = \sum_{m=0}^{M} \hat{f}_m(x) \tag{14}$$

Prediction of price of Al Moammar Information Systems has observed as 106.94523 that is optimum for investing in current bullish market.

Quasi Poisson Regression requires the count variable as represented as dependent variable [22]. It is also known for the generalization of most popular Poisson regression consumed while modeling an over dispersed in form of count variable. Specific variables based on the Poisson Regression process. Author has considered number of events per unit of space and eventually the number of events dependents on the size of the specific events. Table 6 represented the regression results for IT company. Furthermore, author has modeled Y as the specific number of events as Poisson Distribution, as specified in equation 15.

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \tag{15}$$

For y= 0, 1, ∞, λ is the particular mean with respect to space related to interest.

The prominent step in Supervised Learning is fetching the accuracy of the model. The Mean absolute error [14] computes the average of the absolute differentiation between the actual and predicted values in the series of data as formulated in equation 16. MAE evaluates the average of the residuals in series of data calibrated from yahoo finance api.

Table 6. Evaluation of Quasi Poisson Regression.

GEE Regression Results						
	coef	Std error	z	P >|z|	[0.025	0.975]
Intercept	-1.0147	0.147	-6.912	0.000	-1.302	-0.727
Returns	-8.3425	1.567	-5.325	0.000	-11.41	-5.272
Buy, sell, open	0.3501	0.335	1.046	0.295	-0.306	1.006
Open	0.0006	0.001	0.639	0.523	-0.001	0.003
Skew	0.1346		Kurtosis	-1.8786		
Centered Skew	0.1587		Centered kurtosis	-1.5741		

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{16}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{17}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \qquad (18)$$

Mean Squared Error as reprised as the average of the squared dissimilarity across the original and predicted values in the sourced data. It quantifies the variance of specific residuals in large quantities of irregular data as shown in equation 17. Root Mean Squared Error [17] is contemplating as the square root of altogether Mean Squared error as formulated in equation 18. It administers the standard deviation of standard residuals. Fig.11 is concerned with histogram comparison between three errors [19].

Previous studies focused on fundamental aspects of the company thus making it difficult to enter into peaked stage of market with higher chance of downward trend in the future. Future work extends by involving more machine learning methods with multiple companies' implementation with long-term period of investment evolution [17].

Table 7. Comparison analysis of Supervised Learning methods.

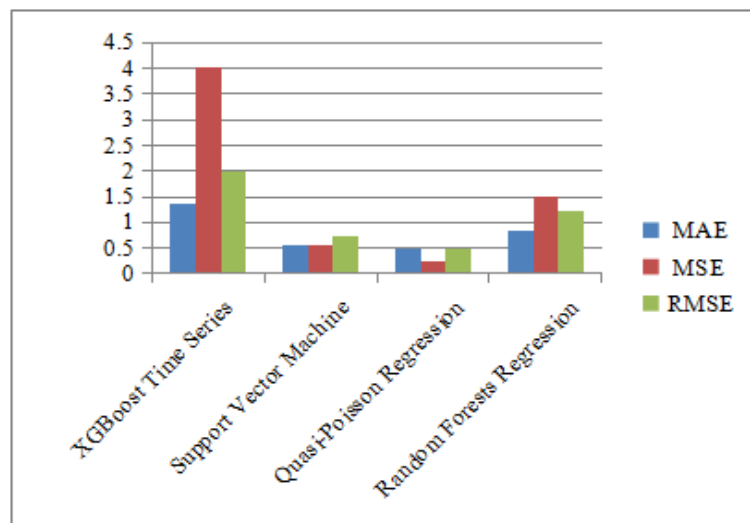| Supervised Learning	MAE	MSE	RMSE
XGBoost Time Series	1.3756160	4.040635	2.0101332
Support Vector Machine	0.56	0.56	0.74833147
Quasi-Poisson Regression	0.474412	0.231796	0.481452327
Random Forests Regression	0.844807	1.50751	1.227807

Fig. 11 Represented of MSE, MAE and RMSE on basis of Supervised Learning.

4. Conclusions

The research also emphasizes on predicting the right price for investing based on Supervised Learning methods involving Support Vector Machine, Random Forest Regression, XGBoost, Auto Arima and Quasi Poisson Regression. Research has found that the right price to investing in this company comes out to be 106.945 on the prediction of previous 6 months period data. Data is sourced though Yahoo Finance api in form of Date, Open, High, Low, Close, Volume, Dividends and Stock Splits. Through this analysis, Xboost Algorithm has outperformed as compare to other supervised

learning methods due to its less complexity and sequence procedure. This solution can be fruitful for newly trained investors who are willing to invest for long-term basis.

Conflicts of Interest: Authors share no conflict of interest in the Research analysis and Prediction of stock market. We confirm that the manuscript has read and approved by all named authors and that are no other persons who satisfy the criteria for authorship but are not listing.

References

[1] K. Zhang, G. Zhong, J. Dong, S. Wang and Y. Wang, "Stock market prediction based on generative adversarial networks", Procedia computer science, vol. 147, pp. 400-406, 2019.

[2] D. Wei, "Prediction of Stock Price Based on LSTM Neural Network", 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), pp. 544-547, 2019.

[3] L. Sayavong, Z. Wu and S. Chalita, "Research on Stock Price Prediction Method Based on Convolutional Neural Network", 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), pp. 173-176, 2019.

[4] H. Abbasimehr, M. Shabani and M. Yousefi, "An optimized model using LSTM network for demand forecasting", Computers & Industrial Engineering, pp. 106435, 2020.

[5] M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques", Procedia Computer Science, vol. 167, pp. 599-606, 2020.

[6] H. R. Patel, A. M. Patel and S. M. Parikh, "Challenges and Applications for Implementing Machine Learning in Computer Vision: Machine Learning Applications and Approaches" in Challenges and Applications for Implementing Machine Learning in Computer Vision, Hershey, PA: IGI Global, pp. 119-135, 2020.

[7] S. Baek, S. K. Mohanty and M. Glambosky, "COVID-19 and stock market volatility: An industry level analysis", Finance Research Letters, vol. 37, pp. 101748, 2020.

[8] 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 (Big Data Service), 2019.

[9] A. F. Wagner, "What the stock market tells us about the post-COVID-19 world", Nature Human Behaviour, vol. 4, no. 5, pp. 440-440, 2020.

[10] D. Shah, H. Isah and F. Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Preddiction Techniques", International Journal of Financial Studies, pp. 2-22, 2019.

[11] N. N. Reddy, V. K. BP and N. E., "Prediction Stock Price Using Sentimental Analysis Through Twitter Data", UTC from IEEE Xplore, 2020.

[12] G. S. Amidwar, R. Kulkarni and M. Muthya, "Stock Prediction Through News Sentiment Analysis", Journal of Architecture & Technology, vol. 9, pp. 36-40, 2019.

[13] G. D. and L. QIN, "Study on the prediction of stock price based on the associated network model of LSTM", International Journal of Machine Learning and Cybernetics 11, vol. 6, pp. 1307-1317, 2020.

[14] T. S. Kumar, "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm", Journal of Artificial Intelligence 2, vol. 3, pp. 185-193, 2020.

[15] F. Feng, H. Chen, X. He, J. Ding, M. Sun and T.-S. Chua, "Enhancing stock movement prediction with adversarial training", Proceedings of the 28th International Joint Conference on Artificial Intelligence, pp. 5843-5849, 2019.

[16] C. Chen, L. Zhao, J. Bian, C. Xing and T.-Y. Liu, "Investment behaviors can tell what inside: Exploring stock intrinsic properties for stock trend prediction", SIGKDD, pp. 2376-2384, 2019.

[17] E. Letizia and F. Lillo, "Corporate payments networks and credit risk rating", EPJ Data Sci., vol. 8, no. 1, pp. 21:1-21:29, 2019.

[18] M. G. Sousa, K. M. Sakiyama, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes and E. T. Matsubara, "BERT for stock market sentiment analysis", ICTAI, pp. 1597-1601, 2019.

[19] C. Li, D. Song and D. Tao, "Multi-task recurrent neural networks and higher-order markov random fields for stock price movement prediction: multi-task rnn and higer-order mrfs for stock price classification", SIGKDD, pp. 1141-1151, 2019.

[20] H. N. Shah, "Prediction of Stock Market Using Artificial Intelligence", 2019 IEEE 5th Int. Conf. Converg. Technol. I2CT 2019, pp. 1-6, 2019.

[21] S. O. Ojo, P. A. Owolawi, M. Mphahlele and J. A. Adisa, "Stock Market Behaviour Prediction using Stacked LSTM Networks-", Proc. - 2019 Int. Multidiscip. Inf. Technol. Eng. Conf. IMITEC 2019, pp. 1-5, 2019.

[22] M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques", Procedia Comput. Sci., vol. 167, no. 2019, pp. 599-606, 2020.

[23] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media news", J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2020.

[24] S. E. Gao, B. S. Lin and C. M. Wang, "Share price trend prediction using CRNN with LSTM structure", Proc. - 2018 Int. Symp. Comput. Consum. Control. IS3C 2018, pp. 10-13, 2019.

[25] A. Nakayama, K. Izumi, H. Sakaji, H. Matsushima, T. Shimada and K. Yamada, "Short-term Stock Price Prediction by Analysis of Order Pattern Images", CIFEr 2019 - IEEE Conf. Comput. Intell. Financ. Eng. Econ., pp. 1-5, 2019.

[26] V. Sharma, R. Khemnar, R. Kumari and B. R. Mohan, "Time series with sentiment analysis for stock price prediction", 2019 2nd Int. Conf. Intell. Commun. Comput. Tech. ICCT 2019, pp. 178-181, 2019.

[27] C. Peng, Z. Yin, X. Wei and A. Zhu, "Stock Price Prediction based on Recurrent Neural Network with Long Short-Term Memory Units", 2019 Int. Conf. Eng. Sci. Ind. Appl. ICESI 2019, pp. 1-5, 2019.