

Stock Market Movement Predictability with Machine Learning Technique: An Evaluation Analysis of Support Vector Machine and Logistic Regression Models

Ernest Kwame Ampomah¹

School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China

Zhiguang Qin²

School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China

Gabriel Nyame³

School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China

Prince Clement Addo⁴

School of Management and Economics
University of Electronic Science and Technology of China
Chengdu, China

Abstract— The stock market has become an integral part of financial system of every nation. It will generally reflect the economic situations of a nation's economy. Predicting movement in the stock market has been of interest to practitioners and researchers from diverse fields. Even though the predictability of the stock market has been challenged, the exist enough evidence in the literature that the stock market can be predicted to some degree. Machine learning technique has become a very popular technique among researchers and practitioners trying to predict the behavior of the stock market. Support vector regression (SVM) and logistic regression (LR) are two widely used machine learning techniques, however, these two techniques are sensitive to scaling, and might not produce a good performance without feature scaling. Hence in this study we compare the performance of SVM and LR with standardization and Min-Max scaling techniques in predicting movement of stock prices. Seven stock data randomly collected from three different stock markets are used in the study. The experiment results indicated that both SVM and LR perform poorly without feature scaling. The performance of SVM and LR improve significantly with both standardization and Min-Max feature scaling. The Kendall's W test is used to ranked the performance of the models using accuracy, F1 score, specificity and AUC metrics. LR_Z-score (logistic regression with standardization scaling) has the highest rank while SVM (without feature scaling) is the least rank. The performance of SVM and LR with standardization scaling is better than their performance with Min-Max scaling.

Keywords— *Machine learning; stock price; support vector machine; logistic regression; feature scaling.*

I. INTRODUCTION

The stock market has become an integral part of financial system of every nation. It will generally reflect the economic situations of a nation's economy. Stock market movements can have a significant economic impact on the macro and micro economy [1]. Predicting movement in the stock market has been of interest to practitioners and researchers from

diverse fields. Forecasting stock price movement is regarded as a difficult task as the market is a non-linear, non-parametric and noisy in nature [2, 3]. Movement in the market is influence by factors such as investors psychology, stock related news, economic conditions, etc. The efficient market hypothesis (EMH) states that it is impossible to consistently achieve risk-adjusted returns higher than the profitability of the overall market as stock prices reflect all information [4]. Contrary to EMH, many studies have concluded that stock market movement can be predicted by using publicly available information, such as past stock data, earnings-price ratios, interest rates, monetary growth rates, inflation rates and dividend yields [5-8]. Currently several approaches exist which help market practitioners to predict market movement. These approaches can be broadly grouped into three. These are fundamental analysis, technical analysis and machine learning technique. Fundamental analysis involves evaluating the intrinsic value of a stock to search for long-term investment opportunities. A fundamental analyst studies the overall economy, industry conditions and management and financial strength of the individual companies [9]. Technical analysis predicts stock price movements by means of historical stock price charts and market statistics without consideration of any underlying economic. It is based on the assumption that if investors are able to find previous patterns in the market, they can generate a fairly accurate forecast of the future price movement [10]. Machine Learning (ML) technique has come about due to advances in computational techniques and information technology. Machine learning is a branch of artificial intelligence (AI) that provides systems the ability to learn and improve automatically from past experience without being programmed explicitly [11]. Prediction of stock market movement with machine learning has gained tremendous popularity because of its ability to handle the non-linear, noisy and dynamic nature of the market better than the other methods. Support vector machine (SVM) and logistic regression (LR) are among the most popular ML algorithms. The SVM has kernel function enables it to

separate the data points in a higher dimension space if they are not linearly separable in a lower dimension [12]. It is memory efficient as it uses a subset of the training samples in the decision function (called support vector). Also, the SVM does not suffer from overfitting, not influence by outliers and has the ability to handle high dimensional data well [13]. Logistic Regression is a very simple machine learning algorithms which is easy to implement yet offers great training efficiency and does not require high computation power. It is very efficient technique which is highly interpretable [14]. A major drawback of SVM and LR is that both of them are sensitive to scaling. Since optimization of SVM happens by minimizing the decision vector, the optimal hyperplane is affected by the scale of the input features, hence, data must be scaled prior to SVM training [15]. LR uses gradient descent as an optimization technique, hence, it requires data to be scaled [16]. This study therefore, aims to comparatively evaluate the performance of both SVM and LR with standardization and Min-Max data scaling techniques in predicting the movement of stock prices.

II. RELATED STUDIES

Several studies have investigated the effectiveness of predicting stock market movement with machine learning techniques. In this section provides a summary of some of these studies. Orimoloye et al, [17] Compared the performance of deep feedforward neural networks and SVM and one-layer NN for predicting stock price indices. They used daily, hourly, minute and tick level data to carry out the study. The results presented indicated that the performance of SVM and one-layer NN was better than DNN when daily and hourly data was used. In contrast, the DNN outperformed the SVM and one-layer NN using minute level data. However, at the tick level, there was not much difference between the performance of the DNN and the shallower architectures. Ismail et al [18], conducted a comparative study of artificial neural network, random forest support, vector machine and logistic regression with persistent homology in predicting the next day direction of movement of Kuala Lumpur Composite Index. The experimental outcome indicated that the performance of support vector machine combined with persistent homology produced the best outcome. Zhou et al, [19] evaluated the performance of SVM in predicting patterns and trends of active and inactive stocks. The authors used multiple heterogeneous data sources to carry out the study. The outcome of the study showed that active stocks produced the highest accuracy when multiple non-traditional data sources are combined, while inactive stocks get the highest accuracy when traditional data sources are combined with non-traditional data sources. Nabipour et al, [20] Compared the performance of eleven machine learning models (which include Decision Tree, Support Vector Classifier, Random Forest, Adaboost, eXtreme Gradient Boosting (XGBoost), Naïve Bayes, k-Nearest Neighbors (k-NN), Logistic Regression and Artificial Neural Network (ANN), Recurrent Neural Network and Long short-term memory) in predicting stock market trends. Technical indicators were used as input, and these technical indicators were applied in two ways. Computing the technical indicators by stock trading values as continuous data, and converting indicators to binary data before using. The experimental results indicated that for the continuous data, RNN and LSTM significantly outperform the other models. In the binary data evaluation, although the deep

learning methods performed better than the other models, the difference becomes less because of the improvement in the performance of the models. Kara et al, [21] compared the performance of artificial neural networks and support vector machines in forecasting movement direction of stock price index. The study found that the artificial neural network model performed significantly better than the SVM model. Ou & Wang, [22] applied ten different machine learning techniques (include Linear discriminant analysis, Quadratic discriminant analysis, K-nearest neighbor, Logit model, Naïve Bayes based on kernel estimation, neural network, Tree based classification, Bayesian classification with Gaussian process, SVM, and Least squares support vector machine (LS-SVM)) to forecast the movement of price of Hang Seng index of Hong Kong stock market. The experiment outcome indicated that that the performance SVM was superior to the other machine learning models.

III. MATERIALS AND METHODS

A. Experimental Design

This study evaluates the performance of SVM and logistic regression (LR) machine learning algorithms in combination with standardization scaling and Min-Max normalization in predicting the movement of stock prices. The study involves six different experiments which are (i) application of SVM algorithm to predict the movement of stock price with feature scaling, (ii) application of SVM with standardization scaling technique (SVM_Z-score) to predict the movement of stock prices, (iii) application of SVM with Min-Max scaling technique (SVM_Min-Max) to predict the movement of stock prices, (iv) application of LR to predict the movement of stock prices without feature scaling, (v) application of LR with standardization scaling technique (LR_Z-score) to predict the movement of stock prices, (vi) application of LR with Min-Max scaling technique (Min-Max_Z-score) to predict the movement of stock prices.

B. Support Vector Machine

SVM is a non-probabilistic binary linear supervised classifier. SVM a kind of linear classifiers which is based on the principle of margin maximization. It performs structural risk minimization, which improves the classifier complexity with the goal of obtaining very good generalization performance [23]. The main goal of SVM classifier is to find a hyperplane (decision boundary) that best separate the two classes during training. The idea behind SVM classification is that the most suitable hyperplane is the one that maximizes the margin between the classes. After computing the hyperplane, new instances are assigned to one of class labels depending on their position relative to the hyperplane. the dimension of the hyperplane is determined by the number of features. The orientation and position of the hyperplane is influence by the support vectors (data points closer to the hyperplane). SVM can be used effectively to do both linear or non-linear classification. However, for non-linear classification task, SVM uses a kernel trick to transform the original feature space into a higher-dimensional feature space based on a user-defined kernel functions such as sigmoid, radial basis function (RBF) and polynomial [24]. Equations 1-3 defines the kernel functions where ϕ is the constant of radial basis and represents

the degree of polynomial function, and are the slope and intercept constant.

$$RBF : K(x, y) = \exp(-\gamma \|x - y\|^2) \quad (1)$$

$$Polynomial : K(x, y) = (\alpha x^T \cdot y + c)^d \quad (2)$$

$$Sigmoid : K(x, y) = \tanh(\alpha x^T y + c) \quad (3)$$

C. Logistic Regression

Logistic regression (LR) is a supervised ML technique which is used to predict the probability that an observation belongs to one of discrete set of target classes. LR converts its output using the logistic sigmoid function to produce a probability value which can then be mapped to the discrete classes [25]. The sigmoid function transforms predicted values into probabilities. Equation 4 define the sigmoid function. An illustration of the sigmoid function is provided by Fig 2. LR has the ability to identify the most effective features used for the classification. The output of the logistic regression is a value between 0 and 1. To map the probability value to a discrete class, a threshold value must be set above which an observation is assigned to one class or another LR is based on the idea that the logarithm of the odds of belonging to a target class is a linear function of the feature vector elements used for task of classification [26]. Mathematically expression of LR is given by equation 5.

$$y = \sigma(x) = \frac{1}{1 - e^{-x}} \quad (4)$$

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_d x_d \quad (5)$$

Where p is the probability of belonging to one class, $\left(\frac{p}{1-p}\right)$ is the odds ratio, and $\alpha, \beta_1, \beta_2, \dots, \beta_d$ are the regression coefficients to be determined based on the data.

D. Standardization Scale

Standardization (Z-score) is a scaling technique that transforms data so that the resulting distribution has the properties of standard normal distribution with mean of zero and a standard deviation of one. The mathematical representation of Z-score is given in equation 6.

$$X' = \frac{X - \mu}{\sigma} \quad (6)$$

μ = mean of the feature values

σ = standard deviation of the feature values

E. Min-Max Normalization

Min-Max normalization (Min-Max) is a scaling technique which transform features so that the resulting data values will be between zero and one. In Min-Max scaling, the least and largest value of each feature is transformed to zero and one respectively. Equation 7 expresses the mathematical representation of Min-Max scaling.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (7)$$

X_{\min} = minimum value of the feature

X_{\max} = maximum value of the feature

F. Research Data

In this study, historical data of seven listed stocks are used and all the data are obtained using yahoo finance application programming interface. The stocks used are randomly selected from three different stock markets (New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations System (NASDAQ), and National Stock Exchange of India Ltd (NSE)). The stock used include Apple Inc (APPL), Abbott Laboratories (ABT), Bank of America Corporation (BAC), Hindustan Petroleum Corporation Limited (HPCL), S&P 500 Index, CarMax, Inc (KMX), and Tata Steel Limited (TATASTEEL). Table 1 gives a description of the stock data used. Forty technical indicators are computed from the open, high, low, close, and volume (OHLCV) variables and used as the input features. Details of technical indicators used as input features is given by table A1-A4 in the appendix section. The study is carried out by splitting each dataset into training and test sets. The opening 70% of each data set is used to train the models and the final 30% of the data set is used as test set. The SVM and LR models is train with the training set and the test set is used evaluated.

TABLE I. DESCRIPTION OF THE STOCK DATA SETS

Data Set	Stock Market	Time Frame	Number of Sample
BAC	NYSE	2005-01-01 to 2019-12-30	3773
ABT	NYSE	2005-01-01 to 2019-12-30	3773
TATASTEEL	NSE	2005-01-01 to 2019-12-30	3278
HCLTECH	NSE	2005-01-01 to 2019-12-30	3476
KMX	NYSE	2005-01-01 to 2019-12-30	3773
MSFT	NASDAQ	2005-01-01 to 2019-12-30	3773
S&P_500	INDEXSP	2005-01-01 to 2019-12-30	3773
XOM	NYSE	2005-01-01 to 2019-12-30	3773
HPCL	NSE	2005-01-01 to 2019-12-30	3476
AAPL	NASDAQ	2005-01-01 to 2019-12-30	3773

IV. EXPERIMENTAL RESULTS

Tables II-V show the experimental results for the accuracy, F1 score, specificity, and AUC metrics respectively for the SVM and LR models. Also, figures 1-4 present bar plots of accuracy, F1 score, specificity, and AUC metrics respectively for the SVM and LR models. It can be observed that SVM and LR perform very poorly on the unscaled stock data sets. However, the performance of the SVM and LR models increase drastically when the data is scaled by either the Z-score or Min-Max techniques. LR_Z-score outperformed the other models on AAPL, ABT, KMX, TATASTEEL and BAC data sets. Similarly, the performance of LR_Min-Max is better than the rest of the models on

S&P_500 and HPCL data sets. Overall, the mean accuracy, F1 score, specificity and AUC of LR_Z-score are the best among the models. Figure 5-11 present the ROC curves of the SVM

and LR models on AAPL, ABT, KMX, S&P_500, TATASTEEL, HPCL, and BAC stock data sets respectively.

TABLE II. ACCURACY VALUES OF THE SVM AND LR MODELS

DataSets	SVM	SVM with Z-Score	SVM with Min-Max	LR	LR with Z-Score	LR with Min-Max
AAPL	0.5342	0.8704	0.8380	0.5528	0.8907	0.8657
ABT	0.5361	0.8796	0.8556	0.5750	0.8879	0.8815
KMX	0.5176	0.8889	0.8398	0.5667	0.8963	0.8833
S&P_500	0.5472	0.8509	0.8407	0.4704	0.8509	0.8676
TATASTEEL	0.5343	0.8959	0.8412	0.5569	0.9088	0.8863
HPCL	0.5136	0.8920	0.8658	0.5520	0.8860	0.8991
BAC	0.5056	0.8509	0.8259	0.5379	0.8676	0.8583
Mean	0.5269	0.8755	0.8439	0.5445	0.8840	0.8774

TABLE III. F1 SCORES OF THE SVM AND LR MODELS

DataSets	SVM	SVM with Z-Score	SVM with Min-Max	LR	LR with Z-Score	LR with Min-Max
AAPL	0.6964	0.8768	0.8480	0.6968	0.8959	0.8736
ABT	0.6980	0.8885	0.8648	0.6681	0.8981	0.8902
KMX	0.6821	0.8915	0.8420	0.5916	0.8976	0.8863
S&P_500	0.7074	0.8553	0.8510	0.1227	0.8574	0.8731
TATASTEEL	0.5929	0.8984	0.8439	0.5720	0.9099	0.8875
HPCL	0.6782	0.8986	0.8702	0.6294	0.8908	0.9020
BAC	0.5997	0.8609	0.8348	0.6407	0.8729	0.8640
Mean	0.6650	0.8814	0.8507	0.5602	0.8889	0.8824

TABLE IV. SPECIFICITY RESULTS OF THE SVM AND LR MODELS

DataSets	SVM	SVM with Z-Score	SVM with Min-Max	LR	LR with Z-Score	LR with Min-Max
AAPL	0.0139	0.8787	0.8290	0.0835	0.9026	0.8628
ABT	0.1267	0.8623	0.8483	0.3174	0.8503	0.8643
KMX	0.1410	0.8964	0.8560	0.5240	0.9156	0.8887
S&P_500	0.0889	0.9059	0.8528	0.9570	0.8896	0.9100
TATASTEEL	0.3957	0.8826	0.8348	0.5283	0.9087	0.8870
HPCL	0.1120	0.8489	0.8530	0.3520	0.8634	0.8923
BAC	0.2755	0.7943	0.7868	0.2566	0.8415	0.8321
Mean	0.1648	0.8670	0.8372	0.4313	0.8817	0.8767

TABLE V. AUC VALUES OF THE SVM AND LR MODELS

DataSets	SVM	SVM with Z-Score	SVM with Min-Max	LR	LR with Z-Score	LR with Min-Max
AAPL	0.5693	0.9523	0.9346	0.6015	0.9569	0.9502
ABT	0.4483	0.9593	0.9401	0.6092	0.9644	0.9564
KMX	0.5366	0.9561	0.9351	0.5821	0.9594	0.9526
S&P_500	0.5713	0.9410	0.9410	0.5891	0.9404	0.9498
TATASTEEL	0.5386	0.9693	0.9386	0.5906	0.9729	0.9619
HPCL	0.4712	0.9482	0.9486	0.5534	0.9495	0.9618
BAC	0.5383	0.9460	0.9209	0.5844	0.9513	0.9385
Mean	0.5248	0.9532	0.9370	0.5872	0.9564	0.9530

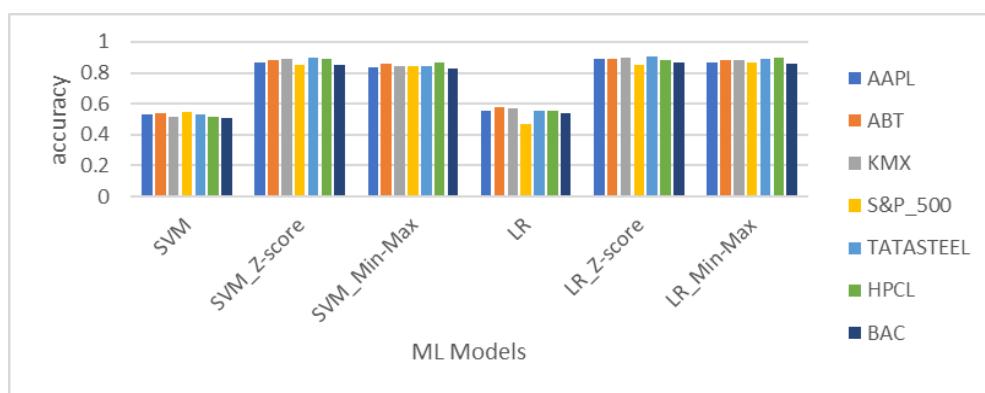


Figure 1: bar chart of the accuracy values of the ML models on the stock data sets

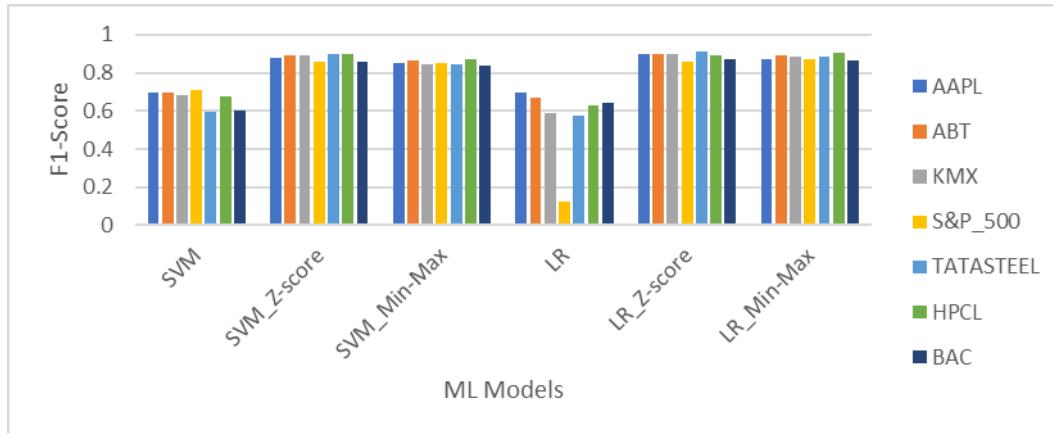


Figure 2: bar chart of the F1-Scores of the ML models on the stock data sets

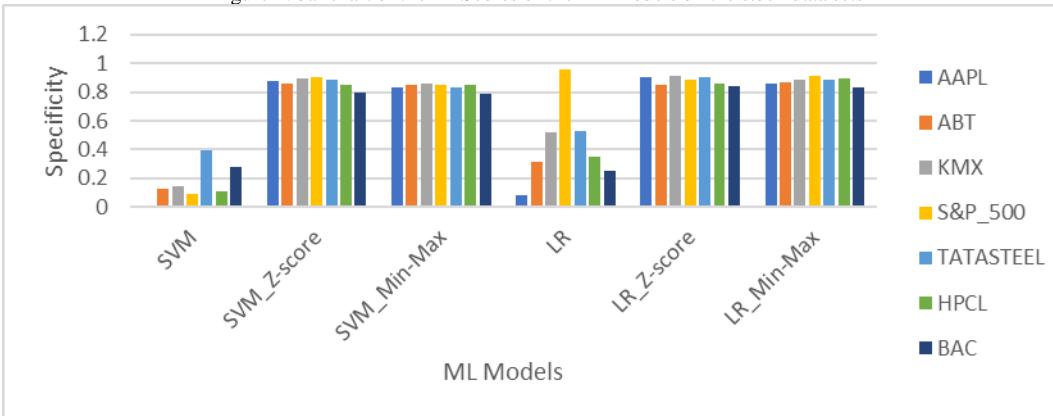


Figure 3: bar chart of the specificity results of the ML models on the stock data sets

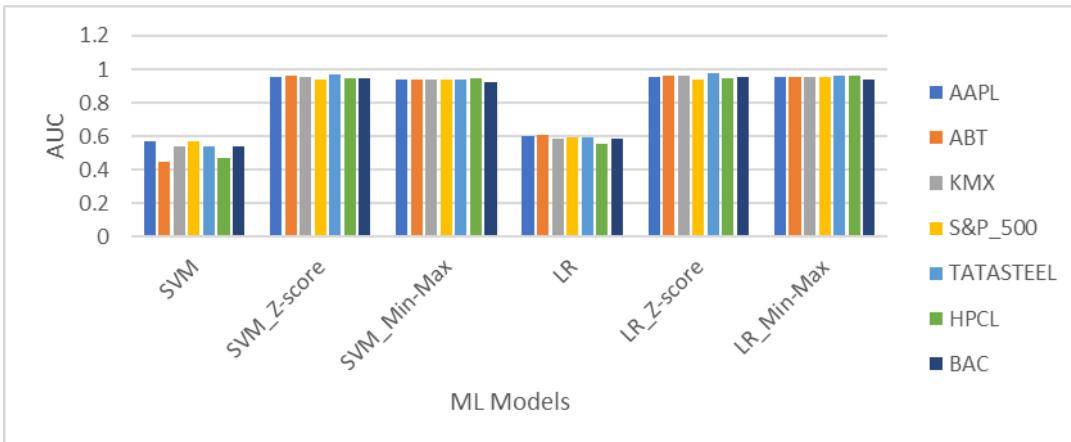


Figure 4: bar chart of the AUC values of the ML models on the stock data sets

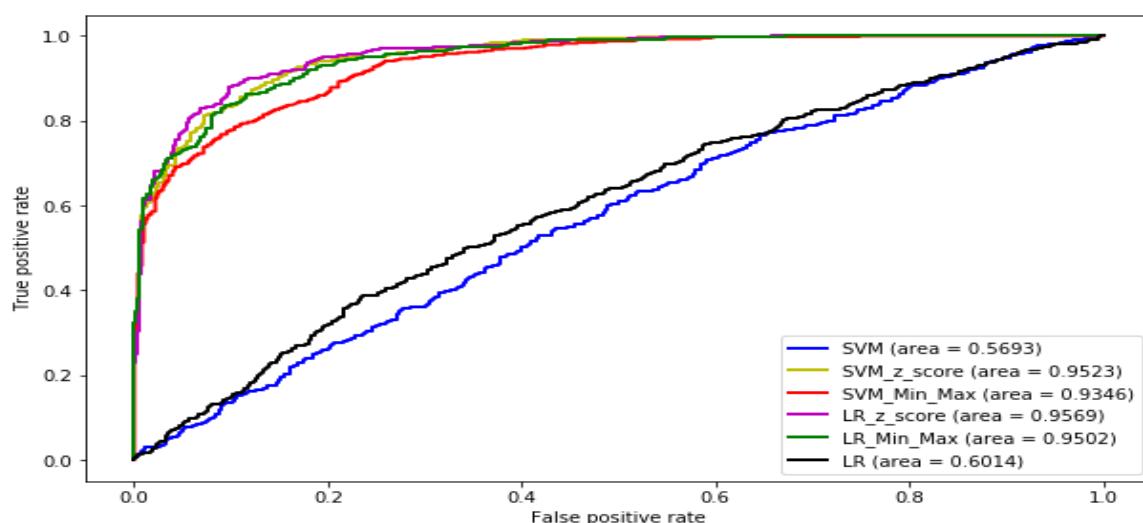


Figure 5: ROC Curves of the SVM and LR models on the AAPL stock data set

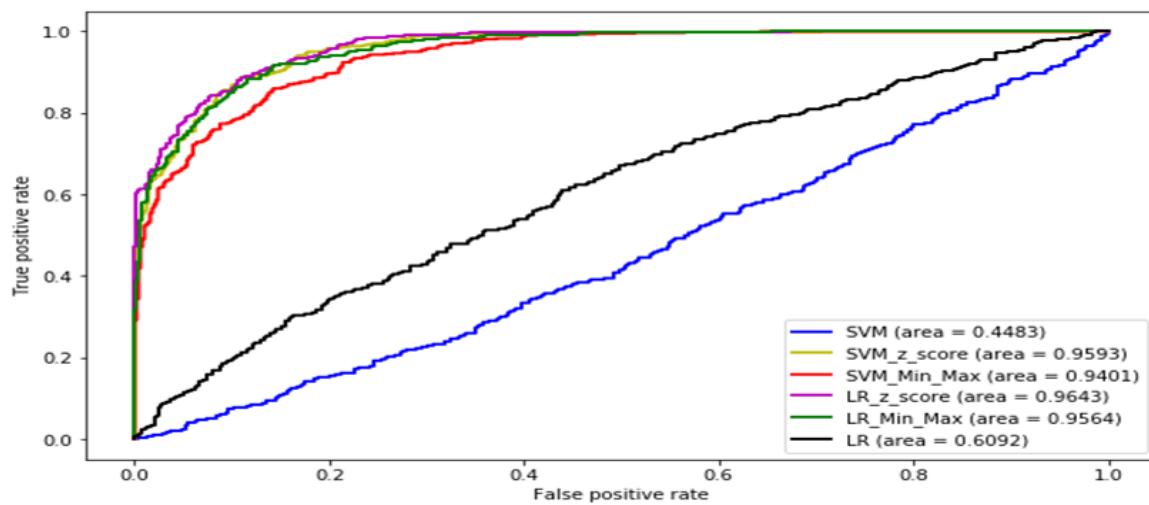


Figure 6: ROC Curves of the SVM and LR models on the ABT stock data set

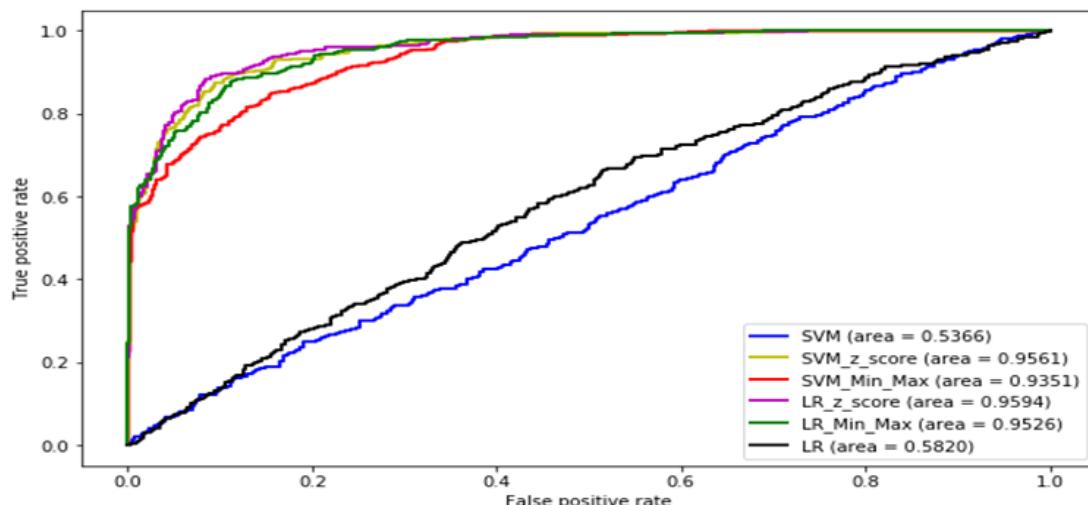


Figure 7: ROC Curves of the SVM and LR models on the KMX stock data set

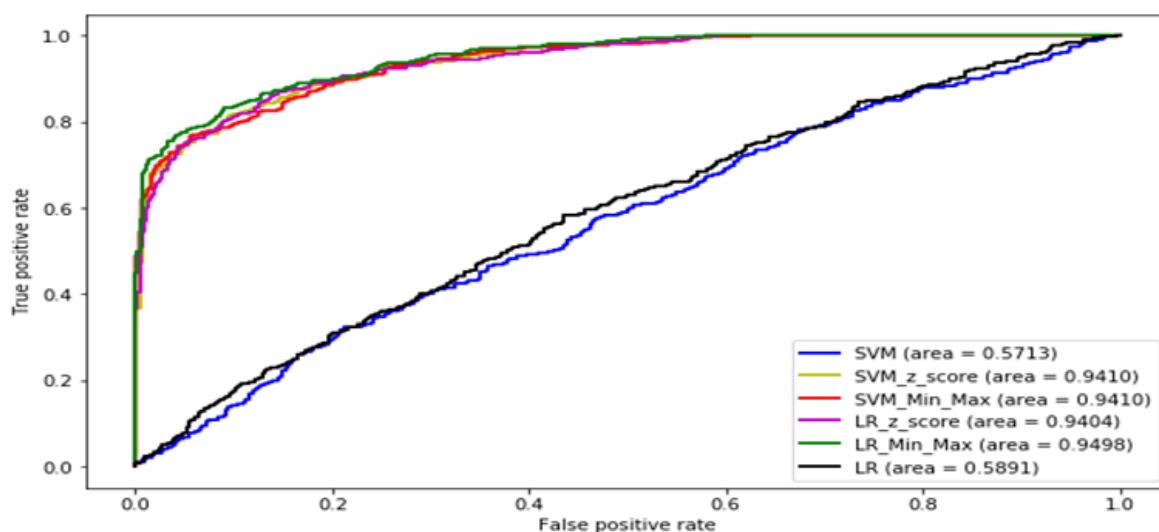


Figure 9: ROC Curves of the SVM and LR models on the S&P_500 Index stock data set

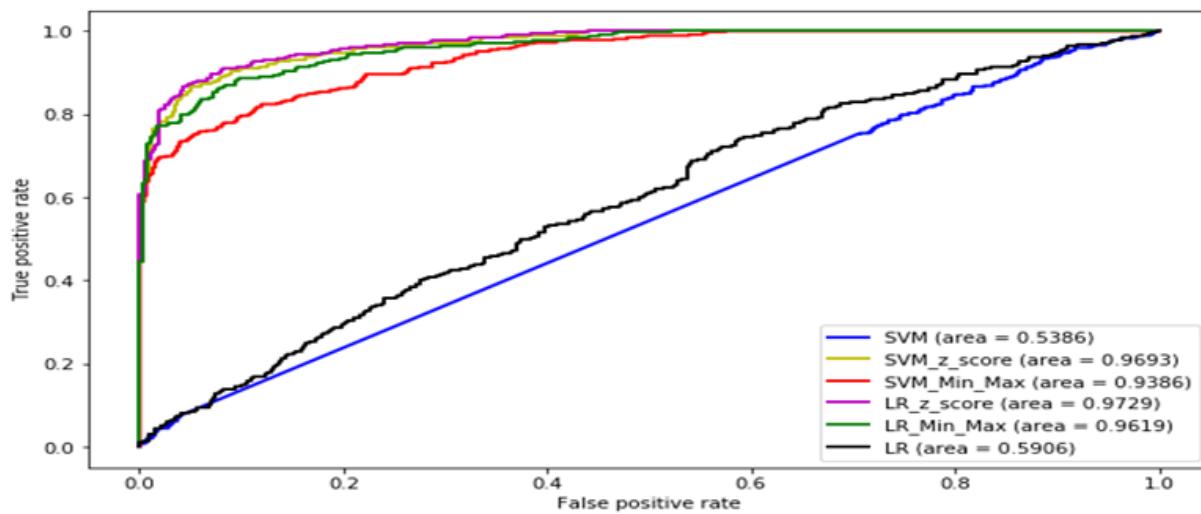


Figure 10: ROC Curves of the SVM and LR models on the TATASTEEL stock data set

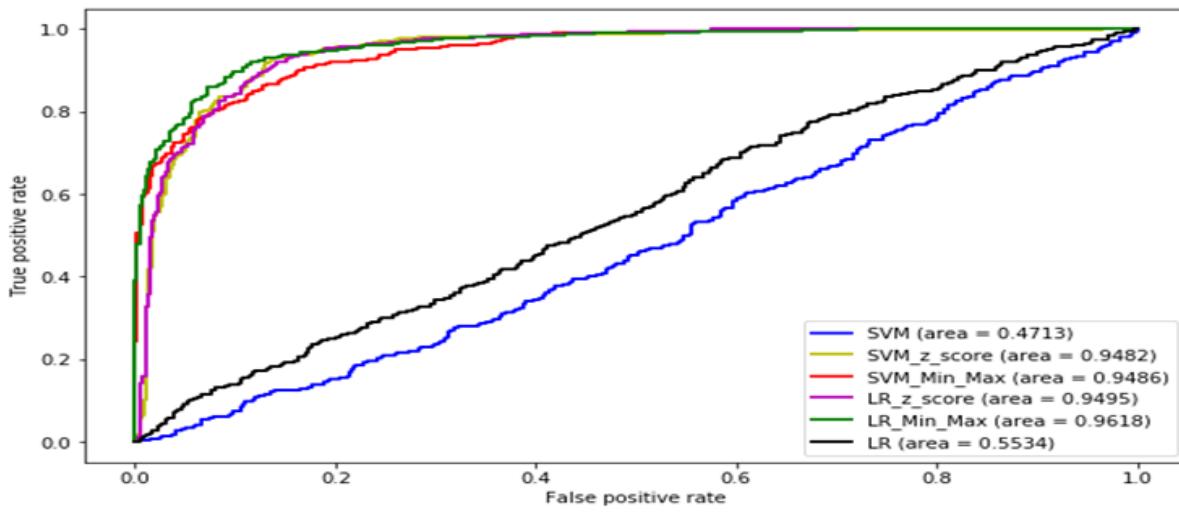


Figure 11: ROC Curves of the SVM and LR models on the HPCL stock data set

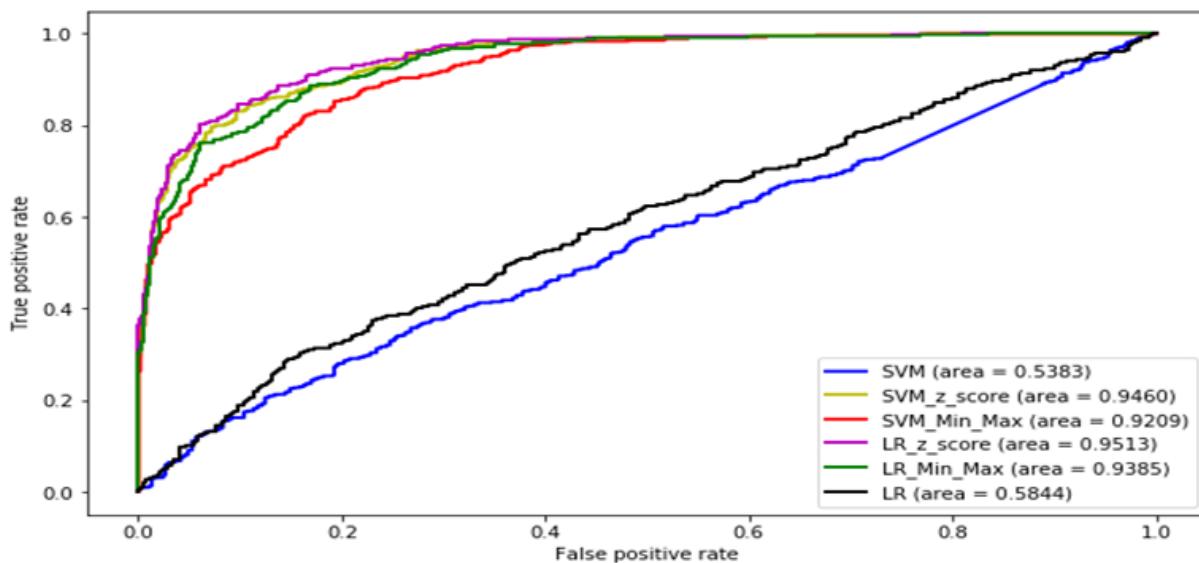


Figure 12: ROC Curves of the SVM and LR models on the BAC stock data set

TABLE VI. RANKINGS OF THE SVM AND LR MODELS BASED ON KENDALL W TEST RESULTS USING EVALUATION METRICS

Measure	W	χ^2	p		Ranks				
Accuracy	0.90	31.30	0.00	Technique	<i>SVM</i>	<i>SVM_Z-Score</i>	<i>SVM_Min_Max</i>	<i>LR</i>	<i>LR_Z-Score</i>
				Mean Rank	1.14	4.64	3.00	1.86	5.50
F1 score	0.89	31.25	0.00	Technique	<i>SVM</i>	<i>SVM_Z-Score</i>	<i>SVM_Min_Max</i>	<i>LR</i>	<i>LR_Z-Score</i>
				Mean Rank	1.71	4.57	3.00	1.29	5.57
Specificity	0.72	25.04	0.00	Technique	<i>SVM</i>	<i>SVM_Z-Score</i>	<i>SVM_Min_Max</i>	<i>LR</i>	<i>LR_Z-Score</i>
				Mean Rank	1.14	4.29	3.00	2.43	5.14
AUC	0.84	29.39	0.00	Technique	<i>SVM</i>	<i>SVM_Z-Score</i>	<i>SVM_Min_Max</i>	<i>LR</i>	<i>LR_Z-Score</i>
				Mean Rank	1.00	4.64	3.36	2.00	5.43

Table VI presents the Kendall W Test ranking of the ML models using the evaluation metrics. For this study a significant level of 0.05 is used, and the Kendall coefficient is deemed significant to provide an overall ranking of the performance of the models when the $p < 0.05$, and $\chi^2 > 11.071$. LR_Z-score and SVM (without feature scaling) are the highest and least ranked respectively. The performance of both SVM and LR with standardization feature scaling are better than their performance with Min-Max feature scaling. The overall ranking using accuracy, F1 score and specificity is **LR_Z-score > LR_Min-Max > SVM_Z-score > SVM_Min-Max > LR > SVM**. However, when using AUC metric SVM_Z-score model is ranked higher than LR_Min-Max model. The overall ranking of the models using AUC is **LR_Z-score > SVM_Z-score > LR_Min-Max > SVM_Min-Max > LR > SVM**.

V. CONCLUSION

This study compares and discusses the effectiveness of predicting stock price movement with SVM and logistic regression machine learning algorithms in combination with standardization and Min-Max scaling techniques. Data of seven randomly selected stocks from three different stock markets are used to carry out the study. Forty technical indicators are computed from the initial stock data and used as input for the machine learning models. The experimental results show that both SVM and logistic regression algorithms performs poorly without scaling of the features. Also, both standardization and Min-Max feature scaling techniques improve the performance of SVM and LR very significantly. The Kendall's coefficient of concordance is used to rank the performance of the models using various evaluation metrics employed in the study. LR_Z-score obtained the highest rank, while SVM (without scaling) recorded the least rank among the models. Both SVM and LR produce better results with standardization scaling than with Min-Max scaling.

Funding: This work was supported by the NSFC-Guangdong Joint Fund (Grant No. U1401257), National Natural Science Foundation of China (Grant Nos. 61300090, 61133016, and 61272527), science and technology plan projects in Sichuan Province (Grant No. 2014JY0172) and the opening project of Guangdong Provincial Key Laboratory of Electronic Information Products Reliability Technology (Grant No. 2013A061401003).

APPENDIX

TABLE VII. VOLUME INDICATORS USED IN THE STUDY

Volume Indicator
Chaikin A/D Line (ADL)
Chaikin A/D Oscillator (ADOSC)
On Balance Volume (OBV)

TABLE VIII. PRICE TRANSFORM FUNCTION USED IN THE STUDY

Price Transform Indicator
Median Price (MEDPRICE)
Typical Price (TYPPRICE)
Weighted Close Price (WCLPRICE)

TABLE IX. OVERLAP STUDIES INDICATORS USED IN THE STUDY

Overlap Studies Indicators
Bollinger Bands (BBANDS)
Weighted Moving Average (WMA)
Exponential Moving Average (EMA)
Double Exponential Moving Average (DEMA)
Kaufman Adaptive Moving Average (KAMA)
MESA Adaptive Moving Average (MAMA)
Midpoint Price over period (MIDPRICE)
Parabolic SAR (SAR)
Simple Moving Average (SMA)
Triple Exponential Moving Average (T3)
Triple Exponential Moving Average (TEMA)
Triangular Moving Average (TRIMA)

TABLE X. MOMENTUM INDICATORS USED IN THE STUDY.

Momentum Indicators
Average Directional Movement Index (ADX)
Average Directional Movement Index Rating (ADXR)
Absolute Price Oscillator (APO)
Aroon
Aroon Oscillator (AROONOSC)
Balance of Power (BOP)
Commodity Channel Index (CCI)
Chande Momentum Oscillator (CMO)
Directional Movement Index (DMI)
Moving Average Convergence /Divergence (MACD)
Money Flow Index (MFI)
Minus Directional Indicator (MINUS_DI)
Momentum (MOM)
Plus Directional Indicator (PLUS_DI)

Log Return
Percentage Price Oscillator (PPO)
Rate of change (ROC)
Relative Strength Index (RSI)
Stochastic (STOCH)
Stochastic Relative Strength Index (STOCHRSI)
Ultimate Oscillator (ULTOSC)
Williams' %R (WILLR)

REFERENCES

- [1] Filis, G. (2010). Macro economy, stock market and oil prices: do meaningful relationships exist among their cyclical fluctuations? *Energy Economics*, 32(4), 877-886.
- [2] Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), 307-319.
- [3] Lai, L. K., & Liu, J. N. (2010, July). Stock forecasting using support vector machine. In 2010 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 1607-1614). IEEE.
- [4] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- [5] Breen, W., L.R. Glosten, and R. Jagannathan. (1990). Predictable variations in stock index returns, *Journal of Finance* 44, 1177-1189.
- [6] Ferson, W.E., and C.R. Harvey. (1993). The risk and predictability of international equity returns, *Review of Financial Studies* 6, 527-566.
- [7] French, K.R., G.S. Schwert, and R.F. Stambaugh. (1987). Expected stock returns and volatility, *Journal of Financial Economics* 19, 3-30.
- [8] Pesaran, M.H., and Timmermann, A. (1994). A generalisation of the nonparametric Henriksson-Merton test of market timing, *Economics Letters* 44, 1-7.
- [9] Abad, C., Thore, S. A., & Laffarga, J. (2004). Fundamental analysis of stocks by two-stage DEA. *Managerial and Decision Economics*, 25(5), 231-241.
- [10] Taylor, M. P., & Allen, H. (1992). The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance*, 11(3), 304-314.
- [11] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
- [12] Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., ... & Olatunji, S. O. (2019). Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. *Computers in biology and medicine*, 109, 101-111.
- [13] Anguita, D., Ridella, S., Rivieccio, F., & Zunino, R. (2003). Hyperparameter design criteria for support vector classifiers. *Neurocomputing*, 55(1-2), 109-134.
- [14] Wang, H. (2019). More Efficient Estimation for Logistic Regression with Optimal Subsamples. *Journal of Machine Learning Research*, 20(132), 1-59.
- [15] Chen, F. L., & Li, F. C. (2010). Combination of feature selection approaches with SVM in credit scoring. *Expert systems with applications*, 37(7), 4902-4909.
- [16] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- [17] Orimoloye, L. O., Sung, M. C., Ma, T., & Johnson, J. E. (2020). Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systems with Applications*, 139, 112828.
- [18] Ismail, M. S., Noorani, M. S. M., Ismail, M., Razak, F. A., & Alias, M. A. (2020). Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange. *Applied Soft Computing*, 106422.
- [19] Zhou, Z., Gao, M., Liu, Q., & Xiao, H. (2020). Forecasting stock price movements with multiple data sources: Evidence from stock market in China. *Physica A: Statistical Mechanics and its Applications*, 542, 123389.
- [20] Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning

and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, 8, 150199-150212.

[21] Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert systems with Applications*, 38(5), 5311-5319.

[22] Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science*, 3(12), 28-42.

[23] Adankon M., Cheriet M. (2009) Support Vector Machine. In: Li S.Z., Jain A. (eds) *Encyclopedia of Biometrics*. Springer, Boston, MA.

[24] Amari, S. I., & Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12(6), 783-789.

[25] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

[26] Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic regression*. New York: Springer-Verlag.