CHAPTER IV : RESULTS AND DISCUSSION

4.1 Evaluation Indicators

Table 4.1-1 below shows all results after data analysis was conducted. Best performance of all trading strategies is in boldface. From here it can be seen that deep learning outperform traditional ML in all the directional evaluation indicator as MLP model of deep learning has the highest value for AR, PR, and F1, and the lowest MSE value. However, traditional ML are better in performance evaluation indicators compared to deep learning as XGB of traditional ML has the highest value for WR, ARR, and ASR.

Table 4.1-1: Trading performance of all trading strategy in BURSA.

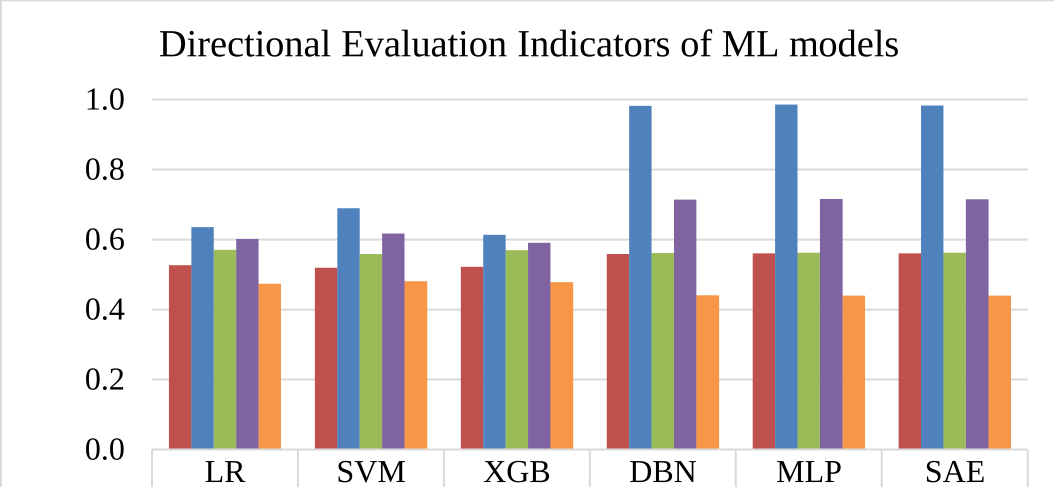
Evaluation Indicators		Index	B&H	**Traditional ML**			**Deep Learning**		
				LR	SVM	XGB	DBN	MLP	SAE
Directional	AR	-	-	0.5263	0.5192	0.5217	0.5589	**0.5605**	0.5603
	PR	-	-	0.6352	0.6892	0.6138	0.9819	**0.9859**	0.9833
	RR	-	-	**0.5708**	0.5587	0.5693	0.5616	0.5623	0.5623
	F1	-	-	0.6013	0.6171	0.5907	0.7145	**0.7161**	0.7155
	AUC	-	-	**0.5108**	0.4950	0.5086	0.4987	0.5000	0.5001
	MSE	-	-	0.4737	0.4808	0.4783	0.4411	**0.4395**	0.4397
Performance	WR	0.5356	0.5144	0.5727	0.5564	**0.5729**	0.5472	0.5476	0.5477
	ARR	0.0615	0.0683	0.2105	0.1158	**0.2223**	0.1682	0.1701	0.1679
	ASR	0.7027	0.3159	1.2597	0.6375	**1.3937**	0.7983	0.8055	0.7982
	MDD	**0.1649**	0.5165	0.2644	0.5165	0.3285	0.5142	0.5079	0.5147

ML: Machine Learning, **AR**: Accuracy Rate, **PR**: Precision Rate, **RR**: Recall Rate, **F1**: F1 Score, **AUC**: Area Under the Curve, **MSE**: Mean Squared Error, **WR**: Winning Rate, **ARR**: Annualized Return Rate, **ASR**: Annualized Sharpe Ratio, **MDD**: Maximum Markdown, **B&H**: Buy & Hold Strategy, **LR**: Logistic Regression, **SVM**: Support Vector Machine, **XGB**: Extreme Gradient Boosting, **DBN**: Deep Belief Network, **MLP**: Multilayer Perception, **SAE**: Stacked Auto-Encoder.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

4.1.1 Directional Evaluation Indicators

Figures 4.1-1 below demonstrates the comparison of directional evaluation indicators on the trading strategies.



	LR	SVM	XGB	DBN	MLP	SAE
AR	0.5263	0.5192	0.5217	0.5589	0.5605	0.5603
PR	0.6352	0.6892	0.6138	0.9819	0.9859	0.9833
RR	0.5708	0.5587	0.5693	0.5616	0.5623	0.5623
F1	0.6013	0.6171	0.5907	0.7145	0.7161	0.7155
MSE	0.4737	0.4808	0.4783	0.4411	0.4395	0.4397

Figure 4.1-1: Comparison of Directional Evaluation Indicators on ML Models.

Based on Figure 4.1-1, it illustrates for AR that MLP has the highest value with 0.5605 and the lowest AR value is SVM with the value 0.5192. It indicates that AR for all the algorithms have the small difference value of 0.04.

Next, the performance for PR are as follows MLP, SAE, DBN, SVM, LR and XGB. There is a big gap between deep learning and traditional algorithms with the value of 0.37 or 37%. This shows that deep learning focuses on “UP” as compared to “DOWN” as compared to traditional ML algorithms.

However, LR of traditional ML has the best performance in RR with the value of 0.5708. There is a small difference for all algorithms for RR with the value of 0.01. This shows that all ML algorithms behaves similarly in concentrating on “UP” as compared to “DOWN”.

This also indicates that it is difficult to produce an algorithm that has a high PR and RR at one time.

For F1, the performance in F1 is as follows by MLP, SAE, DBN, SVM, LR and XGB. There is a noticeable gap between deep learning and traditional algorithms with the value of 0.13 or 13%. This shows that deep learning has a better harmonic balance of precision and recall as compared to traditional ML algorithms.

Next, the highest AUC value is LR and the lowest AUC value is XGB. The highest and the lowest are both from traditional ML algorithms. The values for AUC show a small difference with a value of 0.02. This indicates that all the ML algorithms have no discrimination in distinguishing the classes.

Last of directional evaluation indicator is the MSE. A low MSE implies it has a better performance in regression. The lowest MSE value is MLP with the value of 0.4395 and the highest MSE value is SVM with the value of 0.4808. Thus, it has a difference MSE value of 0.04.

Based on the findings above on directional evaluation indicators, MLP and SAE, both under deep learning, are much more superior to other ML algorithms in terms of predicting the direction of the stock price except for RR and AUC. This is consistent to the results from Soon et. al (2018) as it stated that deep learning has high accuracy. This is due to the multiple of hidden layers that exists in deep neural network. MLP also behaves similarly to SAE and DBN in almost all the evaluation indicators. That would be one of the reason as to why both ML algorithms have work better compared to others.

Therefore, we can determine that deep learning is superior to traditional ML in terms of forecasting the direction of the stock price. This result is consistent with the findings from Chong et al. (2017), Sagir & Sathasivam (2017), Agarwalla et al. (2016), Yiing & Thim (2015) and Monfared & Enke (2014) which revealed that deep learning worked better than traditional

ML in accuracy in stock price prediction. This indicates that Malaysian stocks are compatible with the application of deep learning in forecasting future prices. This is beneficial for investors because can save time in setting the parameter. However, there are other implications as it has less involvement in trading strategy. Another reason that it is good because sometimes there are other influential parameter that is not known or lack in understanding to investors or researchers that they cannot explain. But with deep learning, the algorithms can solve the problem since it is unsupervised algorithm and self-learned even though it is complex and high end (Agarwalla et al., 2016).

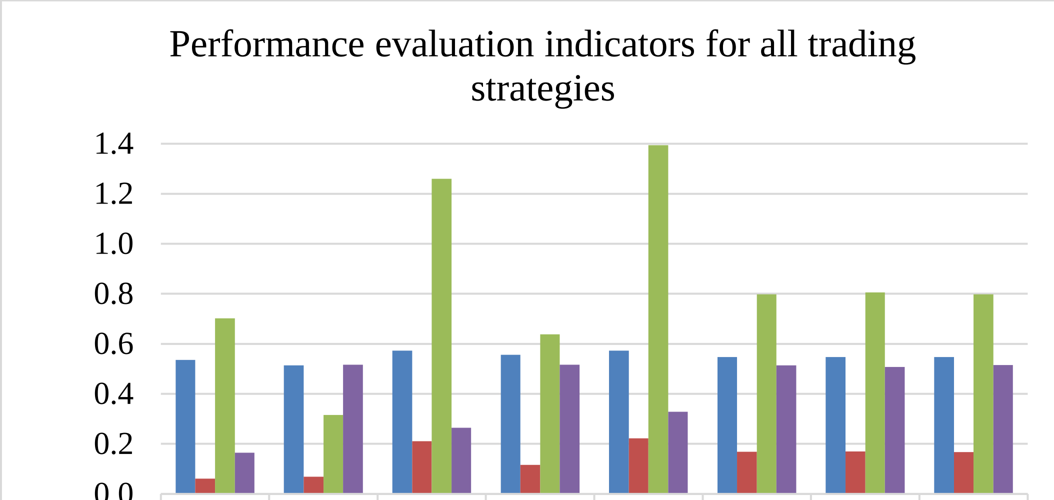
On the other hand, the results are contradicting to the results from Nabipour et al. (2020), Butt et al. (2019), Dongdong et. al (2019), Dey et al. (2016) and Krauss et al. (2016) which exclaimed that traditional ML work better than deep learning. This implies that Malaysian stocks are less compatible with the application of traditional ML in predicting future prices. The investors can use traditional ML in trading strategy to foresee the stock price if they want but it is less effective as compare to deep learning. Possible reasons that traditional ML is less efficient could be because it relies too much on the parameter setting and it is a supervised meaning the investors or researchers have to know to some extent to make it works. If there is a lack of knowledge, the algorithm would not function properly.

Even though the market geography of some past studies is from the same continent as this study and some are from different parts of the world, the impact of market geography on price prediction cannot be made ascertain. This is because previous findings have different features, number of data set, different techniques in data analysis and uses different evaluation indicators. Also, there could be other external influential factors that can affect the price prediction is not included as the parameters for this study.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

4.1.2 Performance Evaluation Indicators

Figures 4.1-2 below demonstrates the comparison of performance evaluation indicators on the trading strategies.



	Index	BAH	LR	SVM	XGB	DBN	MLP	SAE
WR	0.5356	0.5144	0.5727	0.5564	0.5729	0.5472	0.5476	0.5477
ARR	0.0615	0.0683	0.2105	0.1158	0.2223	0.1682	0.1701	0.1679
ASR	0.7027	0.3159	1.2597	0.6375	1.3937	0.7983	0.8055	0.7982
MDD	0.1649	0.5165	0.2644	0.5165	0.3285	0.5142	0.5079	0.5147

Figure 4.1-2: Comparison of Performance Evaluation Indicators.

Based on Figure 4.1-2, it shows that for WR, XGB has the highest value with 0.5729 followed by LR and SVM. Hence, implies traditional ML has better WR values compared to deep learning and other strategies. The higher the value, the better the performance in profitability of the stock.

Next, the highest ARR is XGB with the value of 0.2223 followed by LR with the value of 0.2105 and MLP with the value of 0.1701. The positive the value the better the performance of the trading strategy. All strategies have positive values. But traditional ML algorithms has better ARR values compared to other strategies.

Other than that, the highest ASR is XGB with value of 1.3937 followed by LR with the value of 1.2597. ASR implies the ratio of an extra return from an extra volatility endured for holding a risky asset. Only this two has achieve an acceptable value. Other strategies have value less than value of 1 which means it is less optimal.

Last but not least, MDD represents the risk assessment in the trading strategy. It is favoured to have a low maximum drawdown as it indicates the small losses in the investment. The lowest MDD is from Index with the value of 0.1649 followed by LR with the value of 0.2644 and XGB with the value of 0.3285. Index has the lowest MDD mostly because it is regularly maintained and regulated thus have lower risk as compared to other trading strategies.

Based on the findings above on performance evaluation indicators, XGB and LR, both under traditional ML, are much more enhanced compared to other ML algorithms. Therefore, it can be determined that traditional ML algorithm is superior to deep learning in terms of profitability and risk assessment. This result is consistent with the findings from Krauss et al. (2016) and Dash & Dash (2016). This implies that Malaysian stocks are compatible with the application of traditional ML in trading strategy to increase return and lessen the loss in investment. The investors can use traditional ML in trading strategy.

On the other hand, this result is contradicting with the findings from Yong et al. (2017), M'ng & Aziz (2016) and Yao et al. (1999) which stated that deep learning are more profitable than traditional ML. It is worth to mention the finding from Yong et al. (2017) that shows that DNN produces high profit however yields less accurate results as the model predict further into the future. The finding from this study is the opposite to the findings from Yong et al. (2017). Traditional ML models of this study produces higher profit than deep learning models but yields less accurate results.

We can see that the results gained from this study for performance evaluation indicators for Malaysian market is consistent with the results from United States and India but contradicting to results from Singapore and Malaysia. Even though market geography of the previous studies are in the same region, South East Asia, the impact of market geography on price prediction cannot be made ascertain. This is because past findings apply different features, number of data set, different techniques in data analysis and uses different evaluation indicators. Other than that, there could be other external influential factors that can affect the price prediction is not included as the variables for this study.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

4.2 Statistical Test

The statistical testing called Kruskal-Wallis H Test is performed for all the evaluation indicators, which is AR, PR, RR, F1, AUC, MSE, WR, ARR, ASR, and MDD for all the trading strategies. The results obtain from the hypothesis test of $H_{0,j}$ and $H_{A,j}$ found that the p-value is less than 2.2e-16. Thus, it shows that all the evaluation indicators between all the trading strategies have statistically significant differences. Therefore, to further the analysis of the study on understanding the specific significant difference between the trading strategies, Nemenyi Post-hoc Test is performed as shown in Appendix B. Below are the results obtain for Nemenyi Post-hoc Test for every evaluation indicators.

(1) Table 4.2-1 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on AR of all the ML models. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-1: Pairwise comparison analysis between AR of ML models.

	LR	SVM	XGB	DBN	MLP
SVM	**0.0005**				
XGB	0.1162	0.5750			
DBN	**0.0037**	**0.0000**	**0.0000**		
MLP	**0.0000**	**0.0000**	**0.0000**	0.8023	
SAE	**0.0000**	**0.0000**	**0.0000**	0.8410	1.0000

Based on Table 4.2-1, the AR between traditional ML with deep learning algorithms are statistically significant meaning the AR value shows significant differences. Similarly, LR with SVM show significant difference for AR. Meanwhile, the AR of deep learning with deep learning has insignificant differences. Other than that, XGB has no significance difference for AR with LR and SVM.

(2) Table 4.2-2 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on

PR of all the ML models. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-2: Pairwise comparison analysis between PR of ML models.

	LR	SVM	XGB	DBN	MLP
SVM	0.2239				
XGB	0.2239	**0.0001**			
DBN	**0.0000**	**0.0021**	**0.0000**		
MLP	**0.0000**	**0.0000**	**0.0000**	0.7234	
SAE	**0.0000**	**0.0002**	**0.0000**	0.9918	0.9646

Based on Table 4.2-2, the PR all algorithms are statistically significant except the PR of deep learning with deep learning shows insignificant differences. Likewise, LR has no significance difference for PR with SVM and XGB.

(3) Table 4.2-3 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on RR of all the ML models. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-3: Pairwise comparison analysis between RR of ML models.

	LR	SVM	XGB	DBN	MLP
SVM	**0.0000**				
XGB	0.2239	**0.0000**			
DBN	**0.0000**	**0.0368**	**0.0000**		
MLP	**0.0000**	**0.0004**	**0.0003**	0.8253	
SAE	**0.0000**	**0.0002**	**0.0006**	0.7181	1.0000

Based on Table 4.2-3, the RR all algorithms show statistically significant difference except for LR with XGB and for deep learning with deep learning that shows insignificant differences. MLP and SAE has the highest p-value therefore shows it behaves similarly.

(4) Table 4.2-4 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on F1 of all the ML models. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-4: Pairwise comparison analysis between F1 of ML models.

	LR	SVM	XGB	DBN	MLP
SVM	0.2239				
XGB	0.2239	**0.0001**			
DBN	**0.0000**	**0.0027**	**0.0000**		
MLP	**0.0000**	**0.0000**	**0.0000**	0.7509	
SAE	**0.0000**	**0.0001**	**0.0000**	0.9585	0.9957

Based on Table 4.2-4, the difference of F1 between traditional ML with deep learning algorithms are statistically significant. Similarly, LR with SVM show significant difference for F1. However, the F1 of deep learning with deep learning has insignificant differences. Other than that, LR has no significance difference for F1 with SVM and XGB.

(5) Table 4.2-5 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on AUC of all the ML models. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-5: Pairwise comparison analysis between AUC of ML models.

	LR	SVM	XGB	DBN	MLP
SVM	**0.0000**				
XGB	0.2239	**0.0000**			
DBN	**0.0000**	0.0986	**0.0000**		
MLP	**0.0000**	**0.0017**	**0.0002**	0.8152	
SAE	**0.0000**	**0.0008**	**0.0005**	0.6966	1.0000

Based on Table 4.2-5, the AUC of all algorithms show significant difference except for LR with XGB, SVM with DBN and deep learning with deep learning that shows insignificant differences. MLP and SAE has the highest p-value thus having the highest significant similarities.

(6) Table 4.2-6 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on MSE of all the ML models. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-6: Pairwise comparison analysis between MSE of ML models.

	LR	SVM	XGB	DBN	MLP
SVM	**0.0005**				
XGB	0.1162	0.5750			
DBN	**0.0037**	**0.0000**	**0.0000**		
MLP	**0.0000**	**0.0000**	**0.0000**	0.8023	
SAE	**0.0000**	**0.0000**	**0.0000**	0.8410	1.0000

Based on Table 4.2-6, the MSE between traditional ML with deep learning algorithms are statistically significant. Similarly, LR with SVM show significant difference. Meanwhile, the MSE of deep learning with deep learning has insignificant differences. Other than that, XGB has no significance difference for MSE with LR and SVM.

(7) Table 4.2-7 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on WR of all the trading strategies. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-7: Pairwise comparison analysis between WR of trading strategies.

	Index	B&H	LR	SVM	XGB	DBN	MLP
B&H	0.705						
LR	**0.000**	**0.000**					
SVM	**0.000**	**0.000**	0.697				
XGB	**0.000**	**0.000**	0.705	**0.018**			
DBN	0.064	**0.000**	**0.000**	**0.005**	**0.000**		
MLP	**0.011**	**0.000**	**0.000**	**0.031**	**0.000**	0.999	
SAE	**0.007**	**0.000**	**0.000**	**0.044**	**0.000**	0.998	1.000

Based on Table 4.2-7, almost all the trading strategies show statistically significant difference. However, the WR of deep learning models has insignificant differences with other deep learning models. Similarly, LR has no significance difference for WR with SVM and XGB and also Index with B&H strategy.

(8) Table 4.2-8 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on

ARR of all the trading strategies. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-8: Pairwise comparison analysis between ARR of trading strategies.

	Index	B&H	LR	SVM	XGB	DBN	MLP
B&H	0.705						
LR	**0.000**	**0.000**					
SVM	**0.019**	0.705	**0.000**				
XGB	**0.000**	**0.000**	0.705	**0.000**			
DBN	**0.000**	**0.000**	**0.010**	**0.032**	**0.000**		
MLP	**0.000**	**0.000**	**0.040**	**0.008**	**0.000**	1.000	
SAE	**0.000**	**0.000**	**0.014**	**0.024**	**0.000**	1.000	1.000

Based on Table 4.2-8, the difference of ARR between DBN, SAE and MLP algorithms are statistically insignificant. Similarly, LR with XGB and B&H strategy with Index and SVM shows insignificant difference. Otherwise, all other trading strategies show significant difference.

(9) Table 4.2-9 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on ASR of all the trading strategies. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-9: Pairwise comparison analysis between ASR of trading strategies.

	Index	B&H	LR	SVM	XGB	DBN	MLP
B&H	**0.045**						
LR	**0.000**	**0.000**					
SVM	0.967	0.469	**0.000**				
XGB	**0.000**	**0.000**	0.705	**0.000**			
DBN	**0.014**	**0.000**	**0.010**	**0.000**	**0.000**		
MLP	**0.003**	**0.000**	**0.033**	**0.000**	**0.000**	1.000	
SAE	**0.000**	1.000	**0.000**	0.999	**0.000**	1.000	1.000

Based on Table 4.2-9, most of the trading strategies show statistically significant difference. However, the ASR of deep learning with deep learning has no significant differences. Similarly, SAE has insignificance difference for ASR with B&H strategy, LR with XGB and lastly SVM with Index, B&H strategy, and SAE.

(10) Table 4.2-10 shows results for a pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Distribution approximation for independent samples performed on MDD of all the trading strategies. The p value of two trading strategies with p-value < 0.05 are in boldface.

Table 4.2-10: Pairwise comparison analysis between MDD of trading strategies.

	Index	B&H	LR	SVM	XGB	DBN	MLP
B&H	**0.000**						
LR	0.705	**0.000**					
SVM	**0.000**	1.000	**0.000**				
XGB	**0.019**	**0.000**	0.705	**0.000**			
DBN	**0.000**	1.000	**0.000**	1.000	**0.000**		
MLP	**0.000**	0.979	**0.000**	0.952	**0.000**	0.997	
SAE	**0.000**	1.000	**0.000**	0.999	**0.000**	1.000	1.000

Based on Table 4.2-10, the difference of MDD between deep learning with deep learning algorithms are statistically insignificant. Other than that, SVM and B&H strategy show insignificant difference with all trading strategies except Index, LR and XGB. Index show insignificant difference for MDD for LR. Lastly, XGB show insignificant difference with LR.

UNIVERSITI SAINS ISLAM MALAYSIA
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA