alamSYS: Development of Stock Market Price Forecasting System Using Dynamic Mode Decomposition, Long Short-Term Memory with Arnaud Legoux Moving Average Convergence-Divergence Integration

A Special Problem
Presented to
the Faculty of the Division of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Visayas
Miag-ao, Iloilo

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science by

OLARTE, John Markton M.

Nilo C. Araneta
Adviser

June 2023

Contents

List of Figures

List of Tables

Chapter 1

Results and Discussions

This chapter presents results and discussions from this special problem. Its goal is to provide a comprehensive analysis and interpretation of the data collected for alamSYS's internal and external components. As a result, this chapter is divided into the following sections:

(a) Documentation for alamSYS

(b) DMD-LSTM Results and Discussions

(c) ALMACD Results and Discussions

(d) alamSYS System Tests Results and Discussions

(e) Results and Discussions for the Real World Application of alamSYS

1.1 alamSYS Documentation

The goal of this section is to thoroughly document the current state of the alamSYS in order to facilitate meaningful discussions.

1.1.1 Documentation for alamAPI and Database

xxx

1.1.2 Documentation for alamSYS Preprocessor

xxx

1.1.3 Documentation for alamAPP

xxx

1.1.4 Build and Deployment Guide

xxx

1.2 DMD-LSTM Model Results and Discussions

This section presents and discusses the Deep Learning Model's training, testing, and cross-validation results.

In Table 1.1 the training error metrics are shown for each of the window sizes tested.

Table 1.1: DMD-LSTM Training Error Metrics Scores for Different Window Sizes

Error Metrics	*Window Sizes*			
	5	**10**	**15**	**20**
MSE	0.000037	0.787877	0.006917	0.057851
RMSE	0.006106	0.887624	0.083166	0.240522

Table 1.1 continued from previous page

Error Metrics	*Window Sizes*			
	5	**10**	**15**	**20**
MAE	0.004175	0.755407	0.067645	0.202746
MAPE	**0.000001**	0.000194	0.000017	0.000053

Where it is observed that the best performing model based on having the lowest MAPE score is the DMD-LSTM with a window size of 5. Morever, we can see the differences from each MAPE score for each window size in the Figure 1.1 shown below.

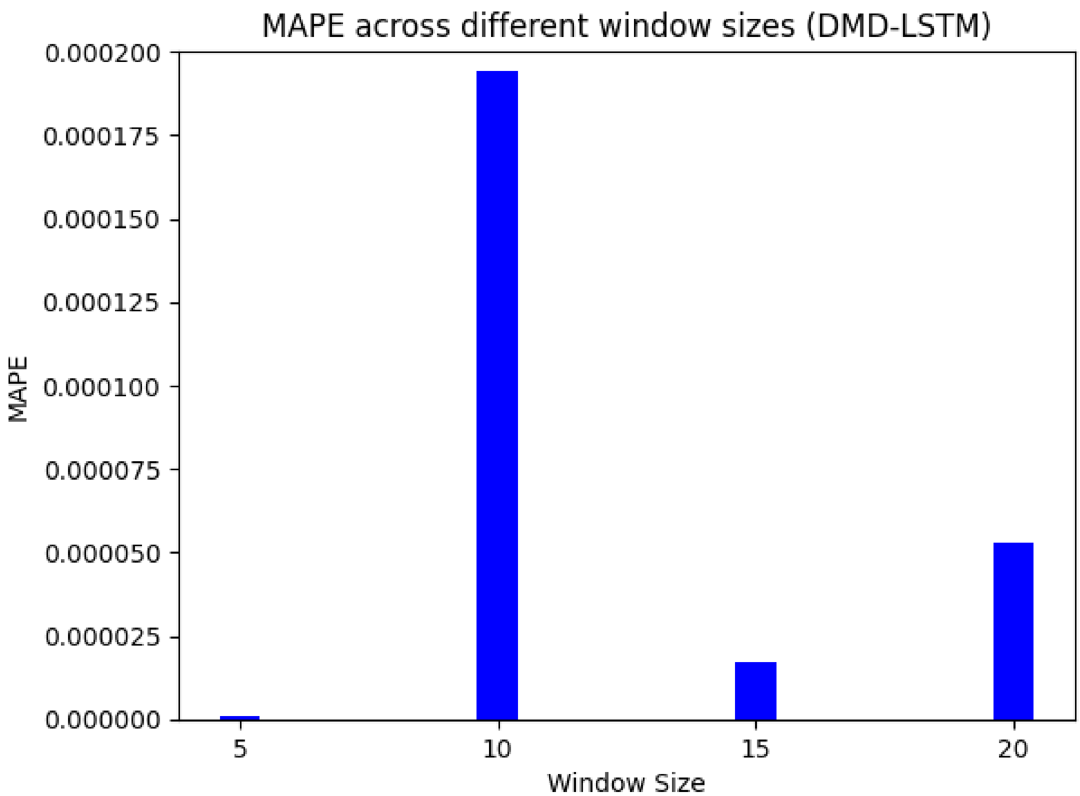

Figure 1.1: Comparison of MAPE Scores for DMD-LSTM Model Training Across Different Window Sizes

The figure above also shows that the MAPE score for window sizes 15 and 20 is higher than the MAPE score for window size 10. MAPE score increases from window size 15 to size 20, indicating that increasing window size may result in a lower performing model.

Furthermore, as previously stated, the window size of 5 results in the best MAPE score being the lowest. Where it outperforms the worst performing model (DMD-LSTM with window size 10) by 0.000193 units. As illustrated clearly in Figure 1.1.

Knowing that the DMD-LSTM model performs as expected based on the training data scores, it is critical that we also examine the training data results from a baseline LSTM. The baseline LSTM is, as the name implies, a simple LSTM model lacking the DMD component. The table below shows the results of the baseline LSTM training.

Table 1.2: Baseline LSTM Training Error Metrics Scores for Different Window Sizes

Error Metrics	*Window Sizes*			
	5	**10**	**15**	**20**
MSE	2912.840703	191.935882	1118.183283	706.136814
RMSE	53.970739	13.854093	33.439248	26.573235
MAE	35.301888	9.480864	22.099720	18.285352
MAPE	0.009618	**0.002527**	0.006024	0.005004

According to the table above, the baseline LSTM with window size 10 performs the best, with the lowest MAPE score of 0.002527 when compared to the other baseline LSTM models.

However, the DMD-LSTM model with window size 5 outperforms it by 0.002526. As a result, the alamSYS makes use of the DMD-LSTM model, specifically the one with a window size of 5. Where from now on, the DMD-LSTM model refers to the DMD-LSTM model with a window size of 5.

Nonetheless, the DMD-LSTM model's performance is limited to the training dataset from PSEI, and it must be cross-validated using data from other stocks, which includes the PSEI validation dataset. The results of this cross-validation is presented in Table 1.3. It should also be noted that cross-validation uses logarithmic normalization as a data preprocessing technique to make the dataset more normal, which aids in analyzing the model's performance with the given dataset. Normalization techniques, in particular, allow for closer variation within the forecasted data. (S.Gopal Krishna Patro, 2015).

Table 1.3: DMD-LSTM Cross-Validation Error Metrics Scores

Stocks	MSE	RMSE	MAE	MAPE
PSEI	0.00002	0.00419	0.00328	1.510000e-03
AC	0.00236	0.04856	0.03414	6.110000e-03
ALI	0.00255	0.05054	0.03645	1.597000e-02
AP	0.00129	0.03596	0.02515	9.220000e-03
BDO	0.00160	0.03999	0.02799	7.250000e-03
BLOOM	0.01883	0.13721	0.06901	1.052898e+12
FGEN	0.00224	0.04733	0.03265	1.197000e-02
GLO	0.00211	0.04595	0.03149	4.680000e-03
ICT	0.00335	0.05785	0.03731	3.005818e+11
JGS	0.00331	0.05752	0.03992	2.009923e+11
LTG	0.01567	0.12518	0.05858	3.583335e+12
MEG	0.00431	0.06565	0.04422	1.393042e+11
MER	0.00326	0.05708	0.03770	9.170000e-03
MPI	0.00273	0.05230	0.03390	2.497000e-02
PGOLD	0.00149	0.03865	0.02818	7.880000e-03
RLC	0.00338	0.05817	0.03978	6.922000e-02
RRHI	0.00131	0.03618	0.02699	6.390000e-03
SMC	0.00137	0.03702	0.02317	5.690000e-03
TEL	0.00178	0.04214	0.03002	4.240000e-03
URC	0.00297	0.05447	0.03742	1.798000e-02

As shown in the table above, the chosen DMD-LSTM model performs well across all other stocks, demonstrating that the model is not overfitted to the training dataset. This score additionally suggests that the model works with non-training data.

The figures below show a 100-day worth of predicted prices versus actual prices to better visualize the performance of the DMD-LSTM model for each stock.

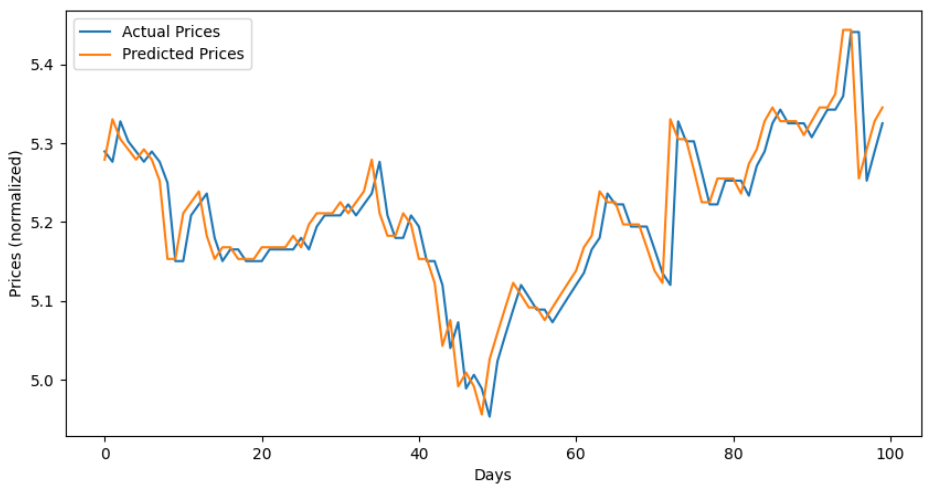

Figure 1.2: Actual vs Predicted Prices on AC for 100 days

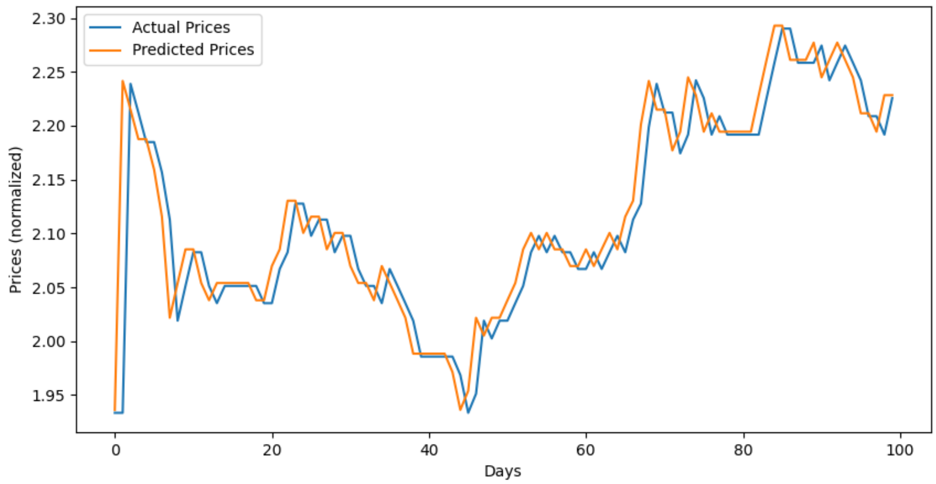

Figure 1.3: Actual vs Predicted Prices for ALI over 100 days

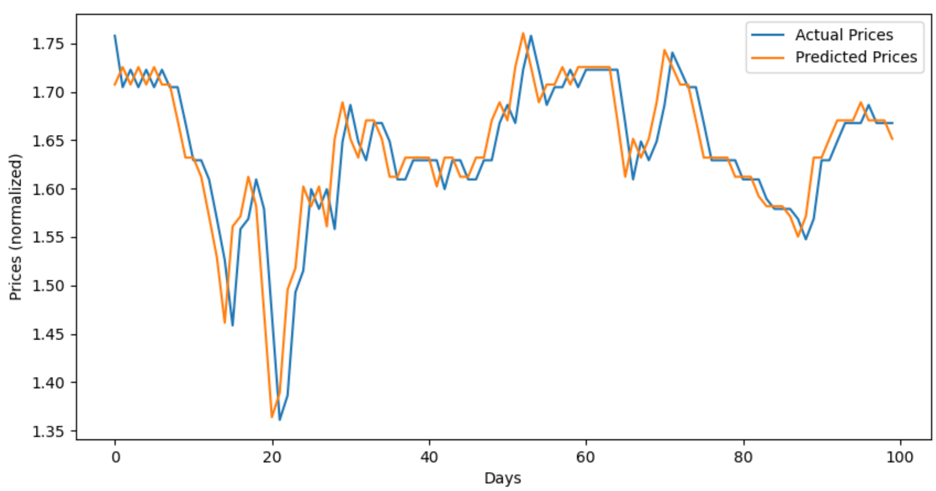

Figure 1.4: Actual vs Predicted Prices for AP over 100 days

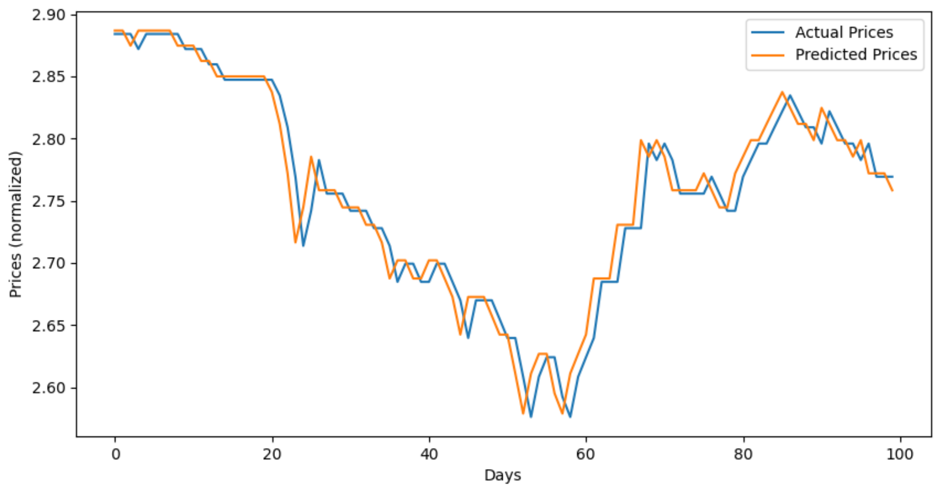

Figure 1.5: Actual vs Predicted Prices for BDO over 100 days

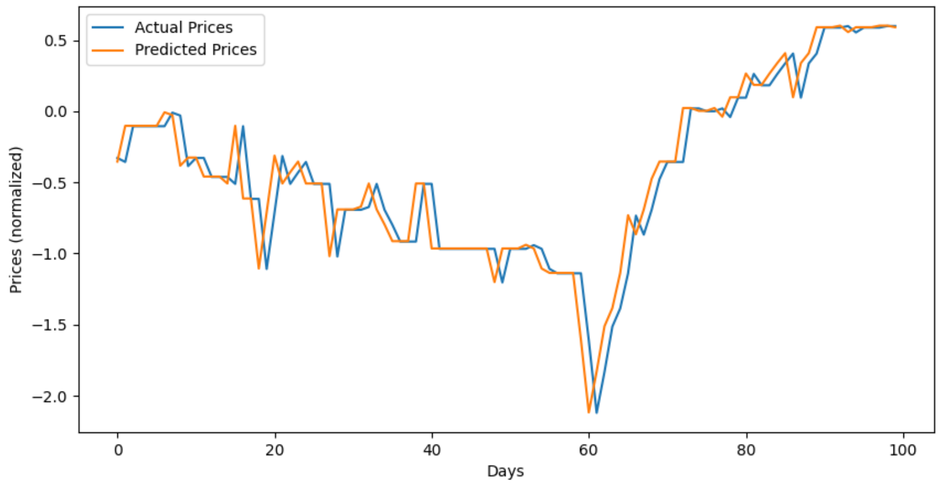

Figure 1.6: Actual vs Predicted Prices for BLOOM over 100 days

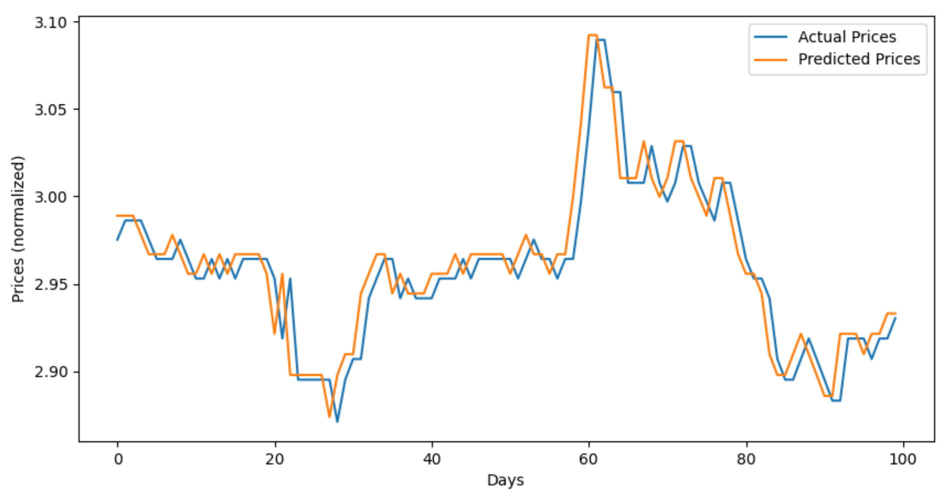

Figure 1.7: Actual vs Predicted Prices for FGEN over 100 days

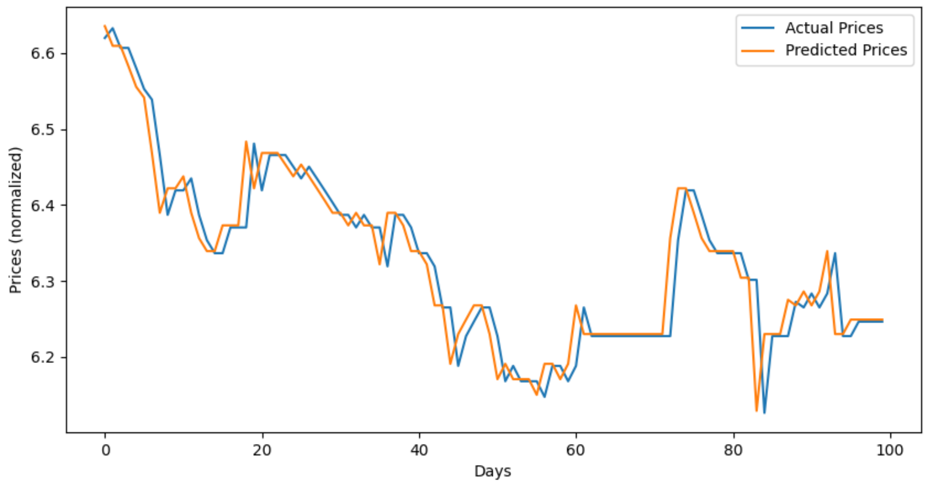

Figure 1.8: Actual vs Predicted Prices for GLO over 100 days

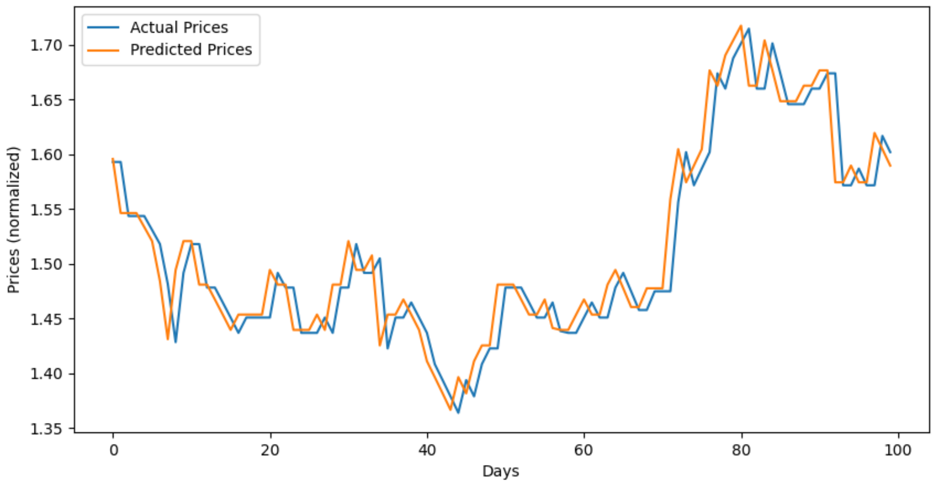

Figure 1.9: Actual vs Predicted Prices for ICT over 100 days

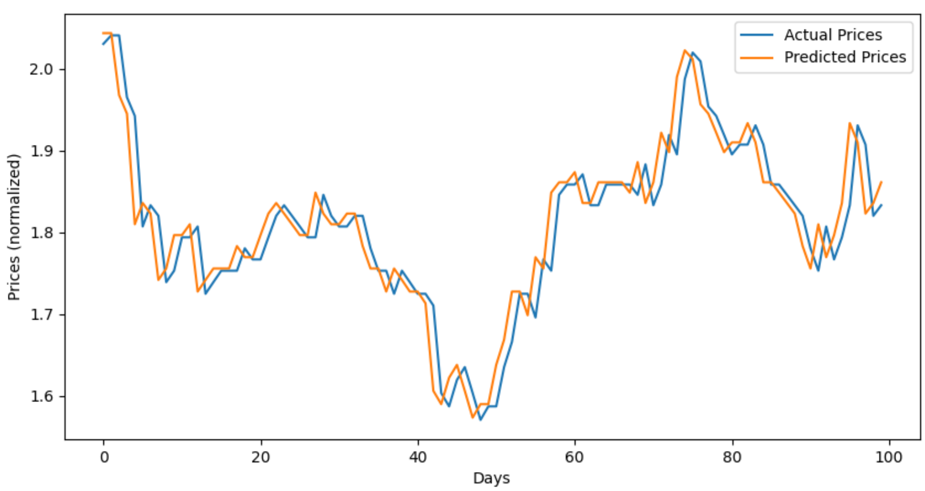

Figure 1.10: Actual vs Predicted Prices on JGS for 100 days

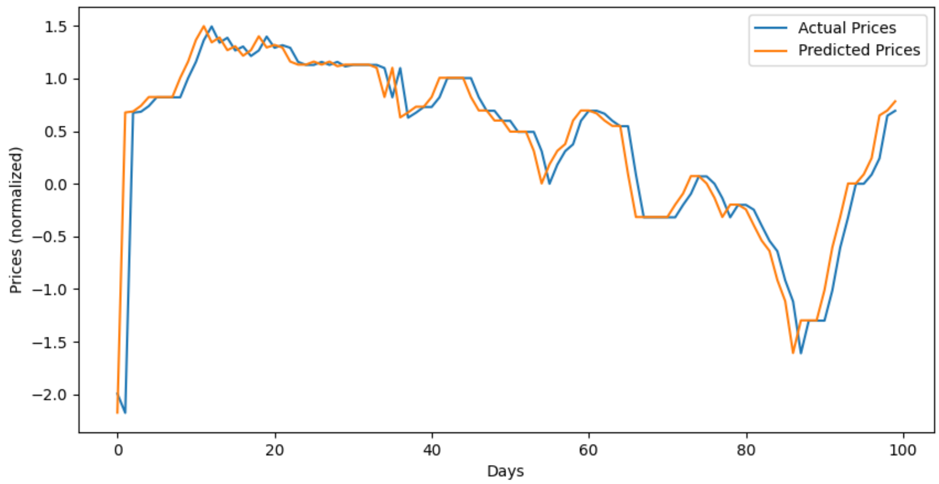

Figure 1.11: Actual vs Predicted Prices on LTG for 100 days

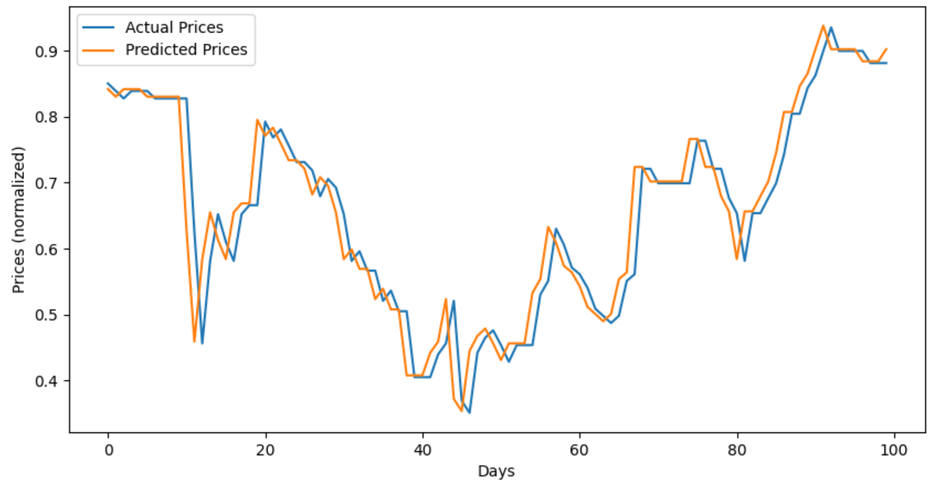

Figure 1.12: Actual vs Predicted Prices on MEG for 100 days

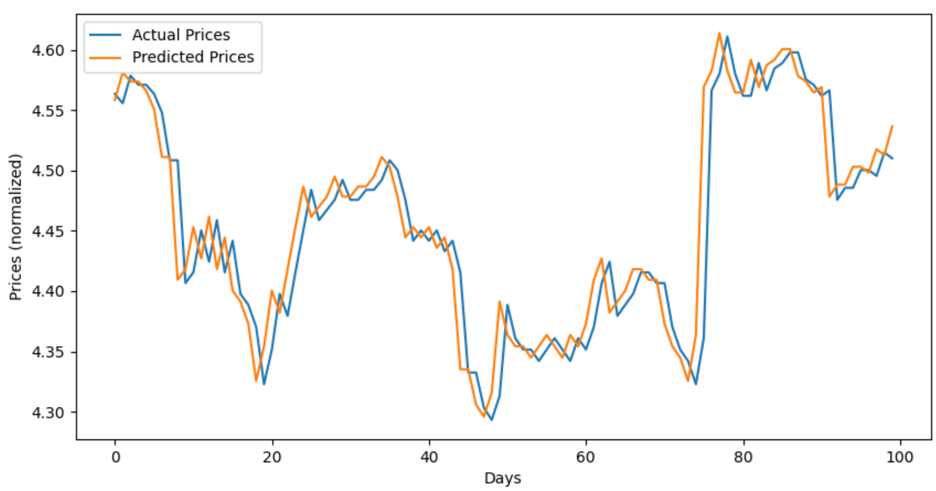

Figure 1.13: Actual vs Predicted Prices on MER for 100 days

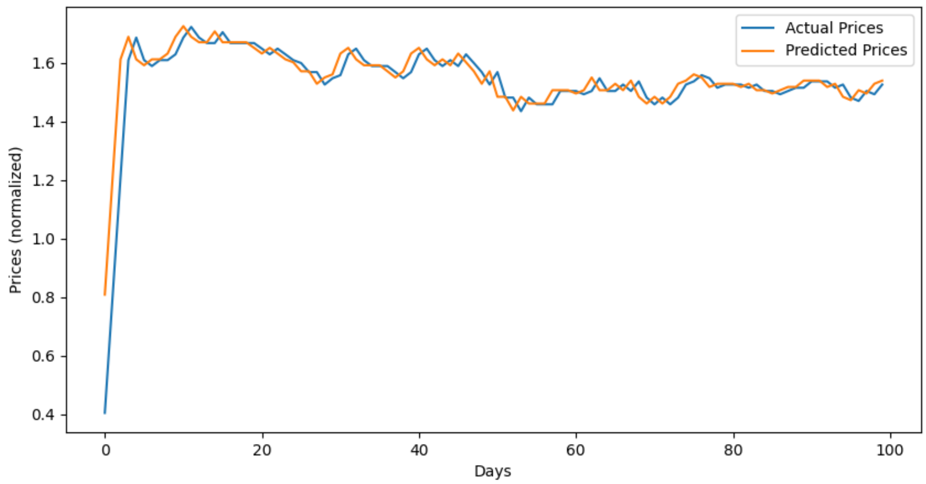

Figure 1.14: Actual vs Predicted Prices on MPI for 100 days

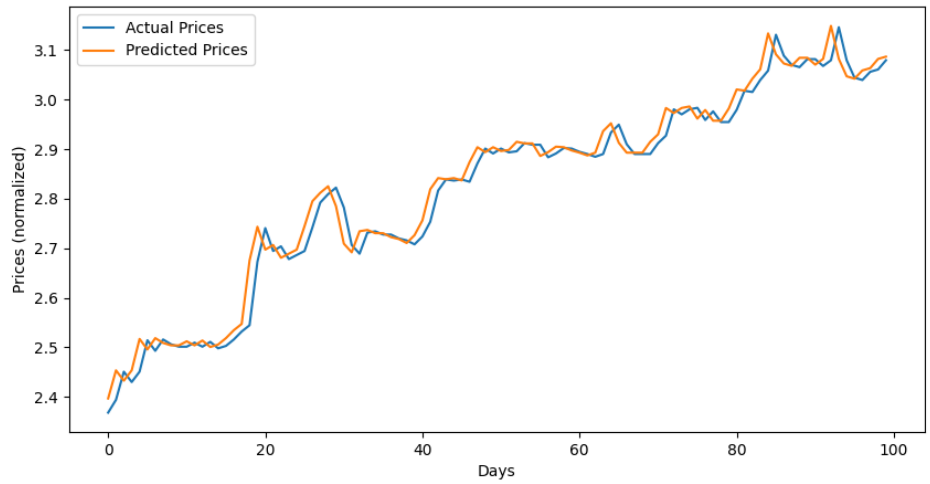

Figure 1.15: Actual vs Predicted Prices on PGOLD for 100 days

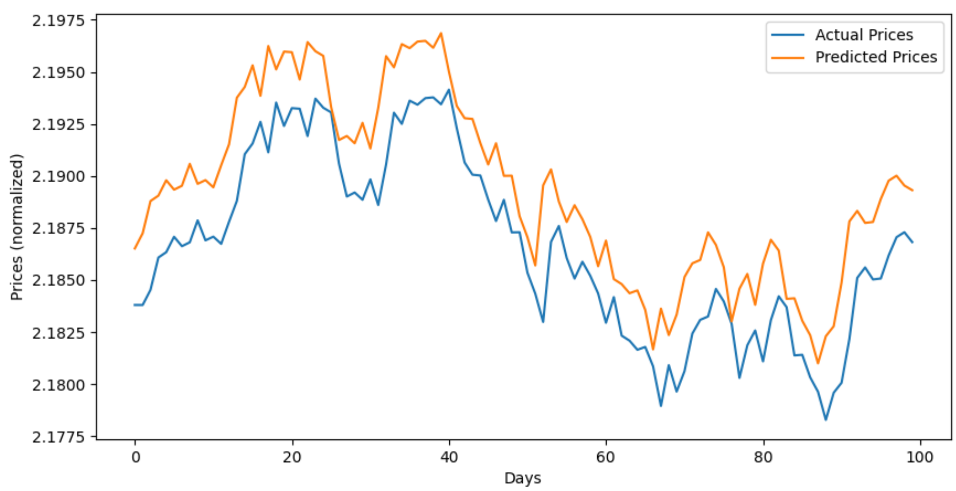

Figure 1.16: Actual vs Predicted Prices on PSEI for 100 days

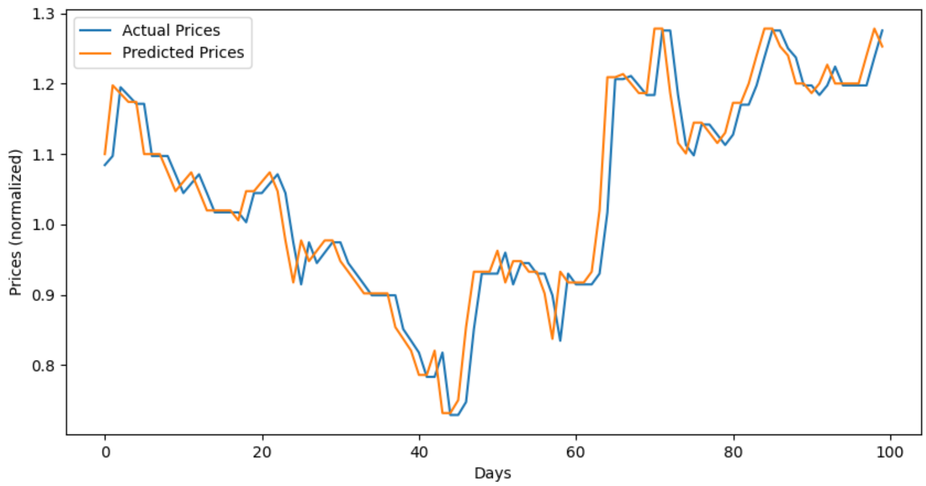

Figure 1.17: Actual vs Predicted Prices on RLC for 100 days

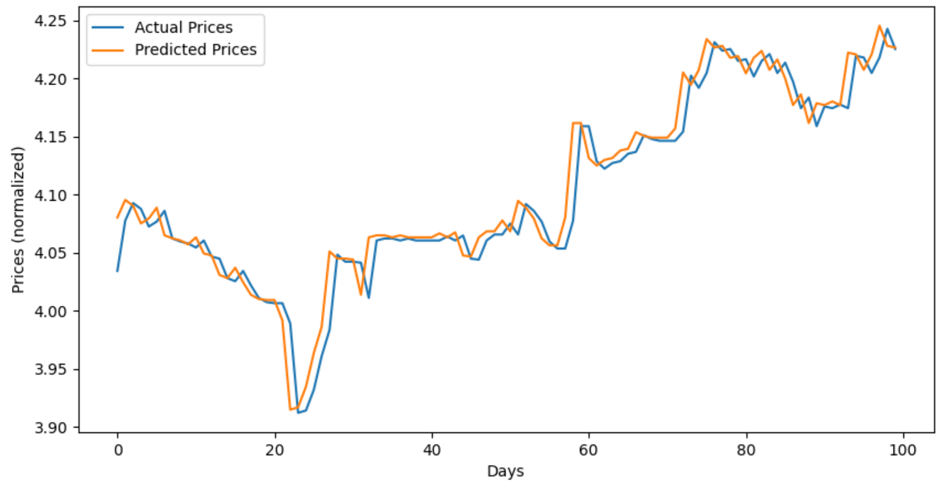

Figure 1.18: Actual vs Predicted Prices on RRHI for 100 days

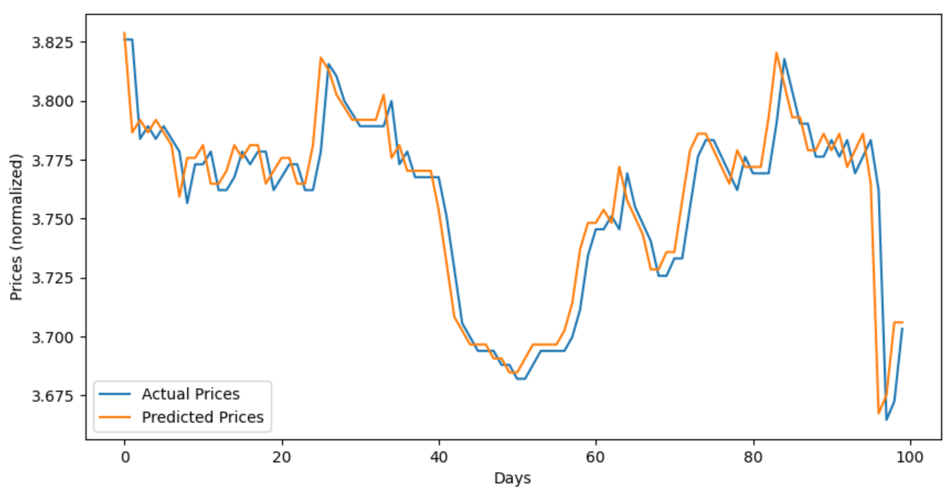

Figure 1.19: Actual vs Predicted Prices on SMC for 100 days

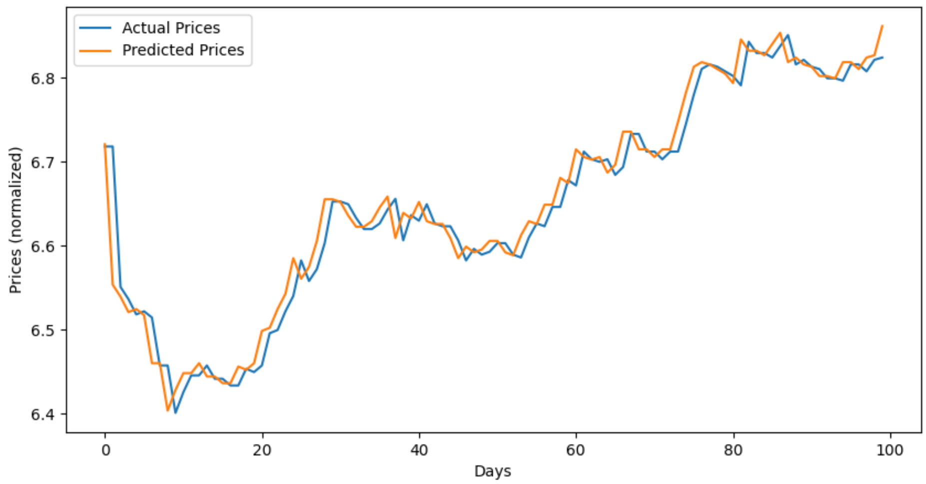

Figure 1.20: Actual vs Predicted Prices on TEL for 100 days

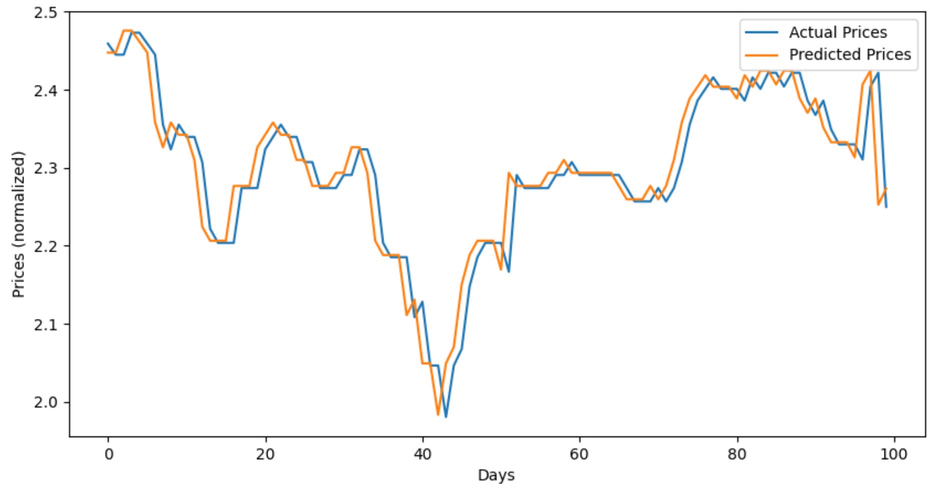

Figure 1.21: Actual vs Predicted Prices on URC for 100 days

The figures above show that the predicted prices follow the actual price trend. In addition, the discrepancy between predicted and actual prices is relatively small, as evidenced by the error metrics scores shown in Table 1.3.

However, the MAPE scores for BLOOM, ICT, JGS, LTG, and MEG range from ten billion to hundred billion. This outlier in the data is, fortunately, just the result of the applied logarithmic normalization, where some of the data in the datasets of the aforementioned stocks are in the negative range, that influence the calculation of the MAPE scores using the scikit-learn library. Because this library handles the calculation of the MAPE scores, there is no way to fix this bug. Moreover, if we take a look at the graphs of the 100 days prediction versus the actual for the aforementioned stocks in Figures 1.6, 1.9, 1.10, 1.11, and 1.12, respectively, it can still be observed that the model performs well on these stocks.

Not to mention that the other error metrics used show the same performance levels across the different stocks when the DMD-LSTM model is utilized. Meanwhile when the data normalization is removed, the MAPE scores for BLOOM, ICT, JGS, LTG, and MEG become 0.068108, 0.037207, 0.039754, 0.057332, and 0.044411 units, respectively.

The successive predictions for the following day and up to ten days were tested using the price data from PSEI in order to make the system's predictions more useful for actual utilization. Table 1.4 shows the MAPE scores for the successive predictions of the DMD-LSTM for each days.

Table 1.4: DMD-LSTM Successive Predictions

Successive Days Predicted	**Actual and Predicted Data Ratio**	**MAPE Score**
1	100%	0.00973
2	80%	0.13403
3	60%	0.15782
4	40%	0.15646
5	20%	0.13910
6	0%	0.12494
7	-20%	0.11283
8	-40%	0.10014
9	-60%	0.08914
10	-100%	0.08976

From the table above xxx

1.3 ALMACD Results and Discussions

xxx

Table 1.5: Optimal Alma Parameters Validation Results

Stock	**Compounded Expected Return**
PSEI	113966.8500
AC	20893.1914
ALI	1072.1418

Table 1.5 continued from previous page

Stock	Compounded Expected Return
AP	690.7100
BDO	2541.9970
BLOOM	495.4600
FGEN	581.0804
GLO	60538.0035
ICT	2815.6103
JGS	1569.8650
LTG	397.2854
MEG	149.2233
MER	8586.0306
MPI	146.0200
PGOLD	721.2700
RLC	649.4767
RRHI	1050.7000
SMC	2557.0770
TEL	72070.5000
URC	3207.5394

xxx

1.4 alamSYS System Tests Results and Discussions

xxx

Table 1.6: Idle System Average Resource Usage Statistics

	alamAPI	alamDB	alamPREPROCESSOR
CPU Utilization (%)	0.168125	0.254313	0.009769

Table 1.6 continued from previous page

	alamAPI	alamDB	alamPREPROCESSOR
Memory Utilization (MiB)	45.718311	166.775377	312.798300

xxx

Table 1.7: Internal Load Average Resource Usage Statistics

	Data Collector	Data Processor	alamSYS PREPROCESSOR (Data Collector & Data Processor)
Failure Rate (%)	0	0	0
Success Rate (%)	100	100	100
Average Runtime (s)	41.72398	8.38061	48.30466
Average CPU Utilization (%)	11.40659	92.71117	20.03138
Average Memory Utilization (MiB)	3.64200	57.09545	794.29436
Average Network Utilization (Mb)	232.73640	154	77.27655

xxx

Table 1.8: Deployment Load Test Results (Buy Requests)

	Number or Requests		
	10	**100**	**1000**

Table 1.8 continued from previous page

	Number or Requests		
Success Rate (%)	100	100	100
Average Processing Time (s)	11.905222	139.618550	1159.773569

xxx

Table 1.9: Deployment Load Test Results (Sell Requests)

	Number or Requests		
	10	**100**	**1000**
Success Rate (%)	100	100	100
Average Processing Time (s)	13.384126	130.119867	1642.995011

xxx

1.5 Results and Discussions for the Real World Application of alamSYS

xxx

Table 1.10: Return Performance Comparison Between alamSYS and PSEI

	Realized Profit (PHP)	Realized Gain (%)
alamSYS	7,839.75	1.51
PSEI	-22,788.90	-13.810

xxx

Chapter 2

Conclusions and Future Work

xxx

References

S.Gopal Krishna Patro, K. K. s. (2015). Normalization: A preprocessing stage. *International Journal of Advanced Research in Science, Engineering and Technology*, *205*. Retrieved from `10.17148/IARJSET.2015.2305`