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

Introduction

1.1 Background and Rationale

The stock market is a type of market that allows businesses to raise capital by selling stock shares to investors. These shares represent a portion of the company's ownership and entitle the holder to a portion of the company's profits as well as voting rights. The stock exchange also serves as a marketplace for investors to buy and sell these shares, allowing for the efficient trading of company ownership. The stock market plays an important role in the growth and development of the economy by allowing companies to raise capital and investors to buy and sell shares (Chen, 2022; The Economic Times, n.d.).

The stock market, contrary to popular belief, is not a form of gambling. It necessitates a significant amount of analytical thinking and risk management, and the returns are determined by supply and demand for a specific stock, rather than false promises or assurances. In other words, rather than being a scam or a gamble, the stock market is a legitimate platform for investing and generating returns (Schwab-Pomerantz, 2021; Adams, 2022; Summers, 2022).

1.1.1 The Philippine Stock Exchange (PSE)

The Philippine Stock Exchange (PSE), Inc. is the country's official stock exchange market. It is a non-stock company founded in 1992 that manages and operates the country's stock market. Individuals who are registered with the PSE can participate in market exchanges (The Philippine Stock Exchange, Inc., n.d.-a).

Furthermore, the Philippine Stock Exchange Index (PSEI) is the main index of the PSE. (PSEI). The PSEI is a market capitalization-weighted price index composed of the PSE's 30 largest and most actively traded companies. These businesses have been pre-selected based on strict criteria such as liquidity and market capitalization. The PSEI is frequently used as a proxy for the overall performance of the Philippine stock market. (Bangko Sentral ng Pilipinas, n.d.) These PSEI companies are often referred to as blue-chip companies because they are typically large, well-established companies with a track record of strong financial performance. In total the PSE has 286 companies listed as of October 2022, offering investors a diverse range of investment opportunities (Fayed, 2022; The Philippine Stock Exchange, Inc., n.d.-b).

1.1.2 Economic Relevance and Benefits of Stock Market Investment

The stock market is widely acknowledged to play an important role in economic growth because it allocates and provides capital to businesses, which drives economic activity and growth. This is evident from the fact that stock market performance is frequently correlated with the gross domestic product (GDP) of the country. (Trade Brains, 2022; Hall, 2022; Bae & Kang, 2017) Furthermore, historical stock price trends can provide insight into broader economic movements (Campbell, 2021).

In a study conducted by Balaba (2017), they discovered that the stock market has a positive impact on the Philippines' economy. The study's findings showed that as the stock market rose, the unemployment rate fell. This is because the performance of the stock market leads to job creation, which in turn leads to economic growth. This, in turn, drives economic growth. This relationship was observed in the Philippines from 2007 to 2017.

1.1.3 Benefits of Investing for the Individual

Individuals in the Philippines can trade shares of publicly traded companies on the Philippine Stock Exchange. Investing in the stock market can provide several advantages to an individual, including:

(a) Protects an individual's money from inflation: Inflation in the Philippines was 6.9% as of September 2022 (Trading Economics, n.d.), while savings account deposit interest rates are only 1-3% annually, (Bureau of the Treasury Bangko Sentral ng Pilipinas, n.d.). This means that savings in deposit banks may not keep pace with inflation, potentially reducing an individual's purchasing power over time. (Royal Bank of Canada Direct Investing Inc., n.d.; EdwardJones, n.d.).

(b) Capital growth opportunities: Investing in the stock market can provide individuals with the opportunity for significant capital growth without the need for direct investment involvement in business operations. This may benefit individuals. Students and working professionals, for example, can increase their capital while remaining focused on their studies or careers. (U.S. Securities and Exchange Commission, n.d.).

1.1.4 Utilization of Machine Learning in Stock Market Trading

In recent years, there has been a surge in interest in the use of machine learning. Learning techniques for predicting stock market movement in the short and long term. As a result, numerous studies and practical applications investigating the use of machine learning in stock market prediction have been conducted.

These efforts aim to improve prediction accuracy and assist investors in making informed decisions (Kumbure, Lohrmann, Luukka, & Porras, 2022; Strader, Rozycki, Root, & Huang, 2020; Soni, Tewari, & Krishnan, 2022; Rea, 2020; Guo, 2022). In this regard, one of the common techniques used is the Long Short-Term Memory (LSTM). LSTM is a deep learning model that is widely used to forecast the stock market. A study by Budiharto (2021) found that LSTM was effective in predicting the Indonesian stock market with 95% accuracy using a short-term data. Which suggests that LSTM can be a useful tool for making short-term stock market forecasts.

The use of Dynamic Mode Decomposition (DMD) for predicting stock market price trends has recently gained traction in the financial industry. DMD is a mathematical method for identifying patterns and trends in large data sets, such as stock market data. It is possible to make more accurate predictions about future stock price movements by applying DMD to stock market data. This can help investors make more informed investment decisions and potentially generate higher returns. However, a study by Lu and Tartakovsky (2020) found that DMD is faster than Proper Orthogonal Decomposition, but it is less accurate.

Other studies have shown that DMD can be effectively applied to the Turkish and Indian stock markets to predict market price trends (Savaș, 2017; Kuttichira, Gopalakrishnan, Menon, & Soman, 2017). These studies show that DMD is simple to implement and can be used as a useful enhancer for making stock market predictions.

1.2 Statement of the Problem

Economic growth in the Philippines is expected to slow in the coming years as a result of the global pandemic, high inflation, and low employment rates (Alegado, Lopez, & Calonzo, 2022; Canto & Romano, 2022; Reuters, 2022).

The lack of free and publicly available stock market predictive systems or tools currently creates a gap in the information available to the public when compared to large private individuals or institutions. These large institutions have the resources to spend a significant amount of money on stock market research, giving them a significant advantage in the investing market. Where, the public is disadvantaged by this lack of access to the same information (Kim, 2022).

Furthermore, the lack of publicly available stock market prediction tools can lead to individuals, particularly first-time investors, making unwise investment decisions, resulting in significant losses and discouragement from investing in the stock market. This is a significant issue because the number of local investors in the Philippine Stock Exchange is already quite small, accounting for only about 1% of the total population. In addition, there has been a significant decline in foreign investment in the Philippines in recent years (Business World, 2022), leading to a corresponding decline in investment volume. As suggested in the study of Balaba (2017), this is expected to have a negative multiplier effect on the country's economic development in the future.

As a result, the creation of a publicly available, simple-to-use, and accurate stock market price trend prediction system could aid in closing the information gap and leveling the playing field for individual investors. This system could help to increase transparency and fairness in the stock market by providing the public with timely and reliable information, resulting in more informed and confident investing decisions and, ultimately, a more stable and prosperous market. Furthermore, such a system could help to increase individual investor participation in the market, resulting in a more diverse and stable market overall. (Statista Research Department, 2022; Commission on Population and Development, 2021).

However, despite the clear and functional benefits of investing in the stock market, many Filipinos remain hesitant to do so for the following reasons:

(a) The difficulties that come with learning the fundamentals of effective stock investing.

(b) The time-consuming nature of technical and fundamental analysis, especially for students and working people on a tight schedule; and

(c) The increased financial risk associated with stock market volatility, as well as the potential for emotional decision-making to jeopardize investments.

These factors *(along with other external and internal factors not listed above)* contribute to a lack of confidence and understanding among potential investors, making it difficult for them to take advantage of the opportunities offered by the stock market.

As such the development of this system, aims to address the following:

(a) The lack of free and publicly available stock market prediction systems or tools.

(b) The time and resources required to study complex traditional market analysis tools, such as fundamental and technical analysis.

(c) The potential for inaccurate market decisions leading to significant investment losses; and

(d) The hesitancy of the Filipino public to begin investing in the Philippine stock market.

1.3 Significance of the Study

The significance of this particular problem lies in the developed system to greatly benefit the stock market, individual investors, and the economy as a whole. Contributions of the system to data-driven investing, financial protection and management, and economic development could provide a valuable resource for investors while also promoting financial stability and growth. Furthermore, the creation of publicly accessible data-driven investing tools or systems may enable more Filipinos to participate in the market and take control of their own financial future. Overall, this special problem has the potential to have a significant impact on the Philippine stock market and economy.

Specifically, this study is significant for the following reasons:

(a) The development of the alamSYS aims to provide the following benefits to the Filipino people:

1. Access to simplified yet accurate information – The proposed system could provide Filipino investors with fast, accurate, and relevant information necessary for effective decision making in the stock market. Using a deep learning model such as LSTM, the system could provide users with the two most important pieces of information: which stocks to buy, and which stocks to sell. This simplified investing model could help investors to make informed decisions and navigate the stock market with confidence.
2. Provide an application interface to facilitate data-driven market decisions – The system could provide users with an intuitive and user-friendly application interface to facilitate data-driven investment decisions, particularly during times when the market is unpredictable or experiencing a downturn. Whereas traditional market analysis tools may not be sufficient to navigate these challenging conditions, the system's forecasting model could provide investors with the insights and guidance they need to make informed and wise decisions. Which would help to promote confidence and stability in the market, even during times of uncertainty.
3. A platform for accessible stock market investment – The system aims to provide all investors, regardless of their investment knowledge, educational attainment, and societal status, with a platform for participating in the stock market. By offering a simplified yet accurate model for investment decision making, the system could empower users to make informed decisions and invest with confidence. This could help to democratize access to the stock market and promote financial inclusion for all Filipinos.

(b) The development of the alamSYS, aims to provide the following benefits to the future developers or researchers:

1. Extension of functionality to other financial markets – The system can be easily adapted or expanded to address related problems in other financial markets, such as investing in government bonds or personal finance management. This flexibility and versatility could make the system a valuable tool for a wide range of investment and financial management scenarios.
2. Testing of new trading algorithms and other machine learning models – The system provides a platform for introducing and testing new data-driven trading algorithms and machine learning models. This could allow future researchers and developers to continually improve the system and keep it at the forefront of data-driven investing technology.
3. Development of a graphical user interface – To further improve the public accessibility of the system, a user-friendly graphical user interface can be developed as a web or mobile application. This could make the system easy to use and intuitive for all users, regardless of their technical expertise.

(c) The development of the alamSYS could help to stimulate economic recovery and development in the country by increasing the number of local investors. As discussed in previous sections, the benefits of the system could encourage more people to invest in the stock market, leading to a multiplier effect that could benefit the economy in several ways. For instance, the increased participation in the market could lead to the creation of jobs and a lowering of unemployment rates. Additionally, the influx of capital into the market could drive fast developments and innovations in various industries. Finally, the increased consumer spending that results from successful investing, stimulates economic growth as well. Overall, the development of the alamSYS could have a positive and far-reaching impact on the economy of the Philippines.

1.4 Objectives

This special problem aims to develop a system that makes investing easier, more publicly available, data-driven, and more approachable by minimizing both the time required for stock price trend analysis, and potential financial risk by using DMD-LSTM and integrate Arnaud Legoux Moving Average Convergence-Divergence (ALMACD) as a trading algorithm. More specifically, it aims to do following:

(a) Develop a Data Preprocessor. Which includes a Data Collector Module (DCM), which collects the end-of-day historical data of a stock from Mondays to Fridays. The data collected is then processed by the Data Processor Module (DPM), which applies the deep learning model and integrate the trading algorithm to the data. Finally, the processed data is given to the Database Updater Module (DUM).

(b) Develop a RESTful API, referred to as alamSYS, using the combination of Python's FastAPI and MongoDB for API endpoints and database, respectively.

Specifically, this was done by doing the following:

1. Develop the following API endpoints:

   1.1 `Home` – This API endpoint outputs a welcome message. Which should inform the user that they have successfully connected to the alamAPI.

   1.2 `Stocks to Buy` – This API endpoint outputs a list of suggested stocks to buy based from the current market price and the predicted price up-trend.

   1.3 `Stocks to Sell` – This API endpoint outputs a list of suggested stocks to sell based from the current market price and the predicted price down-trend.

   1.4 `Stocks Info` - This API endpoint outputs a list of stocks included in the alamSYS and their corresponding information.

1.5 `ML Model Info` - This API endpoint outputs a list of the Machine Learning Models used in the alamSYS and their corresponding information.

1.6 `Stocks Risks Info` - This API endpoint outputs a list of the stocks included in the alamSYS and their corresponding risks values based on value at risk (%), volatility (%), and drawdown (%).

2. Develop a database that stores the results provided by the DPM, and other essential data such as stock information, deep learning model information, and stock risks information about the stock market that is needed to be provided.

(c) Develop a Stock Market Price Trend Forecasting Deep Learning Models by utilizing the dynamic modes in DMD as an additional input parameter to an LSTM model. Afterwards, integrate the forecasting with ALMACD as a trading algorithm and basis for entry and exit positions.

(d) Finally, develop a mobile-based test application, which from hereon maybe referred to as: alamAPP, to showcase the main functionalities of the developed RESTful API. Specifically which stocks to buy and to sell for a given period of time.

1.5 Scope and Limitations

This study was limited only within the companies listed in the Philippine Stock Exchange. Specifically, 20 high volume trade stocks from the year 2021 to 2022 were selected, which are as follows: (1) MEG, (2) JGS, (3) BDO, (4) FGEN, (5) ICT, (6) ALI, (7) SMC, (8) TEL, (9) GLO, (10) BLOOM, (11) RLC, (12) MER, (13) AC, (14) PGOLD, (15) LTG, (16) MPI, (17) AP, (18) RRHI, (19) URC, and (20) PSE Index will be included in the system, instead of the total 286 listed under the Philippine Stock Exchange.

Chapter 2

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

2.1 alamSYS Documentation

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

2.1.1 Documentation for alamAPI and Database

xxx

2.1.2 Documentation for alamSYS Preprocessor

xxx

2.1.3 Documentation for alamAPP

xxx

2.1.4 Build and Deployment Guide

xxx

2.2 alamSYS System Tests Results and Discussions

With the development of alamSYS, we must ensure that all of its components are functioning properly. This section focuses on the system's performance while idle and under load, as well as the API's and database's ability to handle multiple requests at a time.

The Table 2.1 shows the CPU and memory utilization of each of the alamSYS components whenever it is not processing any information. Wherein, the data presented below is gathered by logging the system utilization of alamSYS within an hour.

Table 2.1: Idle System Average Resource Usage Statistics

	alamAPI	**alamDB**	**alamPREPROCESSOR**
CPU Utilization (%)	0.168125	0.254313	0.009769
Memory Utilization (MiB)	45.718311	166.775377	312.798300

From the table above, it is shown that the alamSYS as a whole only utilizes 0.432207% of the total CPU power on average when it is on idle. Wherein, the bulk of the CPU power is being used by the database at 0.254313% which is 58.84% of the total average CPU utilization of the alamSYS.

This result actually shows promising idle performance, as an Ubuntu system's normal idle CPU utilization is less than 10%. However, it should be noted that the alamSYS is not completely idle because background tasks such as scheduling, mongoDB processes, and the API must remain active in order to respond to any API queries.

Furthermore, the lower average CPU utilization for alamAPI and alamPREPROCESSOR could be attributed to the linux distribution used as their base image, as previously discussed on 'Docker-Compose Layer Diagram' of Section **??**.

The CPU utilization for the alamPREPROCESSOR, in particular, was lower than we would have expected given that it is running a schedule checker every second. This indicate that the schedule library utilizes an efficient way to check for schedules. However, it may not be the case for its memory utilization, which is discussed further below.

Moving on, the alamSYS's memory utilization shows that it uses 525.291988 Mebibytes (MiB) or 550.808572 Megabytes (MB) on average. The memory utilization of alamPREPROCESSOR accounts for the majority (59.55%). This demonstrates that, despite having the lowest CPU utilization, the alamPREPROCESSOR consumes more memory than the combination of the alamAPI and alamDB.

Again, as previously discussed, the alamPREPROCESSOR's high average memory utilization is due to the background schedule checking, which in this case is programmed to use more memory than CPU power.

This average utilization result implies that the alamSYS may be able to be deployed on devices with lower specifications. A Raspberry Pi 4, which has a quad core CPU and at least 2GB RAM, is one example (Zwetsloot, 2019). However, idle performance only shows the minimum CPU and memory utilization of the alamSYS components and does not provide a complete picture of its utilization, particularly under load. As a result, we must investigate the system's CPU and memory utilization while under load.

The internal load averages system utilization of the alamSYS' preprocessor is shown in the following table.

Table 2.2: 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

Table 2.2 continued from previous page

	Data Collector	Data Processor	alamSYS PREPROCESSOR (Data Collector & Data Processor)
Average Memory Utilization (MiB)	3.64200	57.09545	794.29436
Average Network Utilization (Mb)	232.73640	154	77.27655

Before exploring into the results shown in Table 2.2, it is critical to first establish a context for how the data was gathered. The data in the above table was collected while the alamSYS, specifically the alamPREPROCESSOR, was subjected to a stress test load of 100 consecutive data collection and processing. Also, the data of average utilization as indicated for each column are independently gathered from each other.

Based on the table above, each component was capable of processing 100 consecutive processes without failure. This means that the alamSYS, specifically the alamPREPROCESSOR, can run at least 100 times in a row without ailing. Also, keep in mind that the alamPREPROCESSOR only collects and processes data once per day, and the developer has included an option for system users or maintainers to manually rerun these processes in cases where the alamPREPROCESSOR fails - for example, due to a lost or slow internet connection, a power outage, and so on.

Meanwhile, the data processor's average runtime is faster than the data collector's, and it runs on an average of 48.30466 seconds. This was already expected because the data collector needed to connect to the internet and was thus constrained by internet speed. Whereas the average speed during the course of this test was 52.98Mbps, this could imply that the data collector's runtime may be slower or faster depending on the internet speed.

Furthermore, the data processor's average runtime was surprisingly fast given that it needed to apply DMD-LSTM to a total of 20 stocks and calculate the position using the ALMACD given a total of 205 data points. Furthermore, looking at its CPU and memory utilization, it can be seen that it uses more than the data collector, with approximately 87.68% more average CPU power used and approximately 93.63% more average memory used.

In line with the CPU and memory utilization, the average CPU utilization of the alamPREPROCESSOR on load is 99.95% higher than when it is idle. And it uses 60.62% more memory on average.

Lastly, looking at the network utilization of the alamPREPROCESSOR, we can see that the data collector used the most network bandwidth, as expected. However, it may be surprising to see that the data processor uses the network when it should not because it does not need to process or collect data from the internet, but this is still within expectations because network bandwidth utilization also accounts for local network utilization. The data processor of alamPREPROCESSOR, in particular, uses the local network to connect to and update the new data in the alamDB.

The following tables show the system's deployment load results. Specifically, to access the buy and sell collections, as shown in Tables 2.3 and 2.4. Additionally, to have a background on the test that results in the following outcomes. It should be noted that the test consisted of three instances of the same tester application running on ten different computers. Where each application requests 10, 100, 1000 buy and sell data from the alamDB via the alamAPI, which is tunneled over the internet for server access outside the university's local network using LocalXpose services.

Table 2.3: Deployment Load Test Results (Buy Requests)

	Number or Requests		
	10	**100**	**1000**
Success Rate (%)	100	100	100
Average Processing Time (s)	11.905222	139.618550	1159.773569

Table 2.4: 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

As per the tables above, the alamSYS was able to handle all consecutive and simultaneous requests from all external devices. It was able to handle 71,000 requests in approximately one hour and 20 minutes.

Furthermore, the data shows that a request is processed in 1.34 seconds on average. Wherein, the fastest processing time from the data collected was 0.93 seconds. This result is within the expected response time of APIs, where good is defined as 0.1 to one second, and any time between one and two seconds is acceptable, as long as it does not exceed two seconds, which users may perceive as an interruption in the process (Juviler, 2022).

It is also worth noting that in a separate internal test for the alamSYS, the alamAPI only takes on average 0.0094 seconds to send its response. And that the additional delay in response time on deployment is caused by network-related factors, such as the time it takes the tunneling service to accept the query and send back the system's response. In fact the network related delay contributes 99.30% of the average deployment response time. Hence, it might be useful to deploy the system in a network that further minimizes this additional delay in the response time.

Meanwhile Table 2.5 shows the average CPU and memory utilization of alamAPI and alamDB as the load testing as stated above, was being processed.

Table 2.5: CPU and Memory Utilization Statistics of alamAPI and alamDB Under Deployment Load Testing

	alamAPI	**alamDB**
CPU Utilization (%)	17.725949	1.070133
Memory Utilization (MiB)	44.620837	129.273305

To further visualize the system utilization over time, Figure 2.1 shows the CPU utilization of alamAPI and alamDB over time. And Figure 2.2 shows the memory utilization of alamAPI and alamDB.

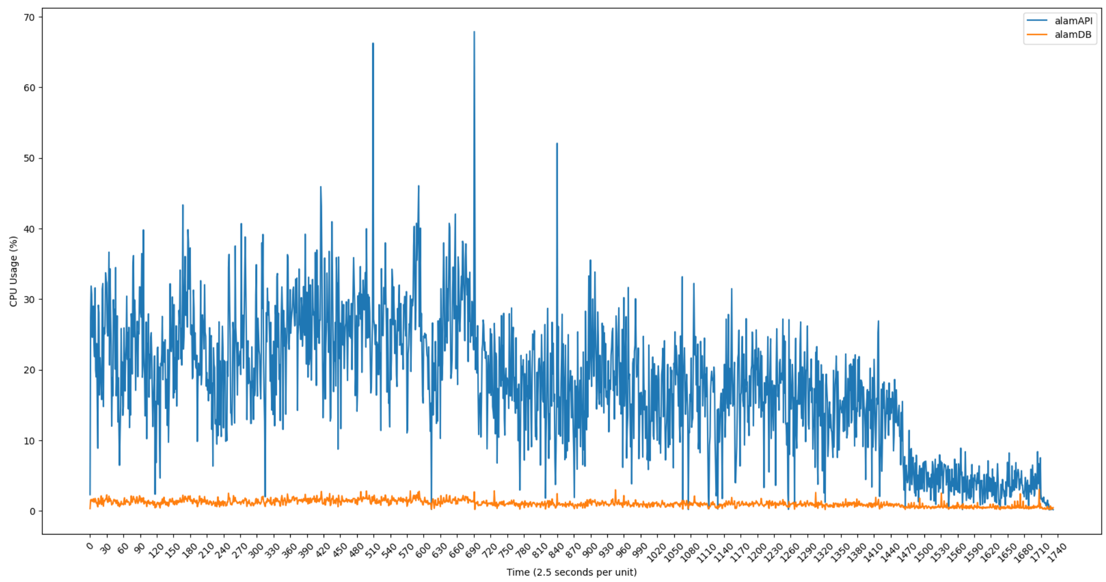

Figure 2.1: Deployment Load CPU Utilization of alamAPI and alamDB Over Time

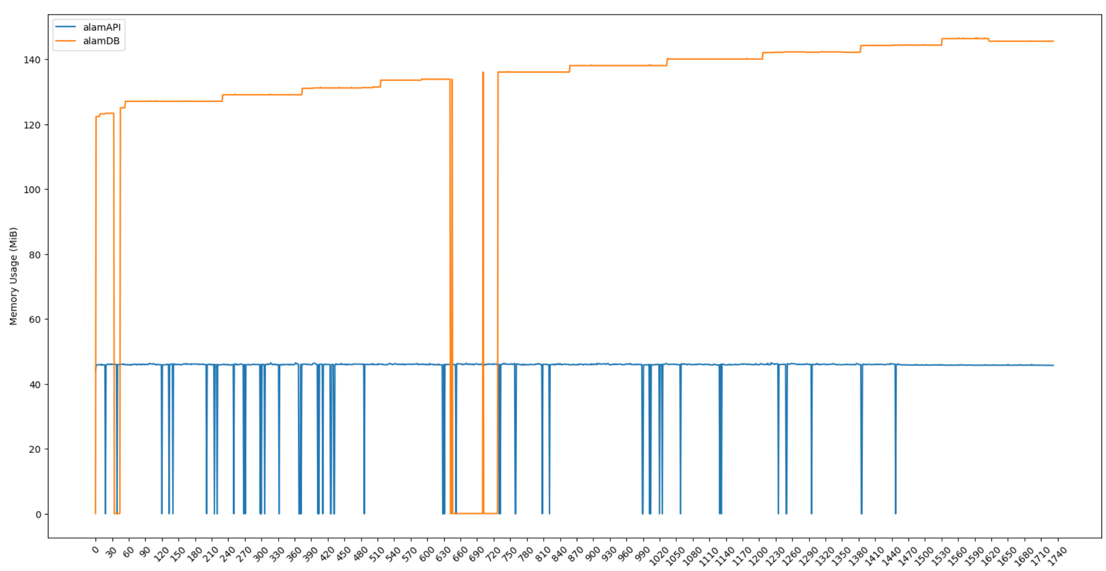

Figure 2.2: Deployment Load Memory Utilization of alamAPI and alamDB Over Time

Based on the table and figures above, processing all requests consumes only 17.73% and 1.07% of the CPU for alamAPI and alamDB, respectively. In terms of memory usage, they are 44.62 MiB and 129.27 MiB, respectively.

Furthermore, at around 1740 time units, the CPU utilization for alamAPI was shown to be reduced. This is due to the fact that most tester applications have already completed processing all of the requests that they are programmed to run, reducing the load on the server by half.

2.3 DMD-LSTM Model Results and Discussions

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

In Table 2.6 the training error metrics are shown for each of the window sizes tested.

Table 2.6: 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
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 2.3 shown below.

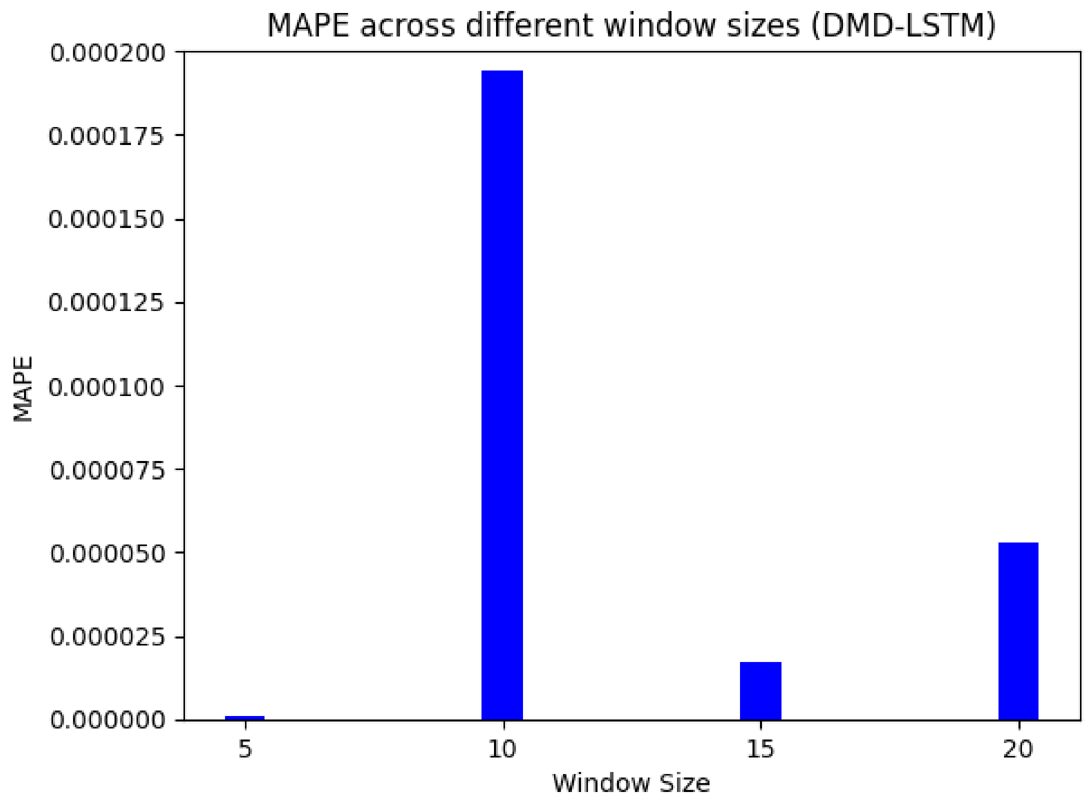

Figure 2.3: Comparison of MAPE Scores for DMD-LSTM Model Training Across Different Window Sizes

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 2.3.

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 2.7: 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 2.8. 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 2.8: 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.

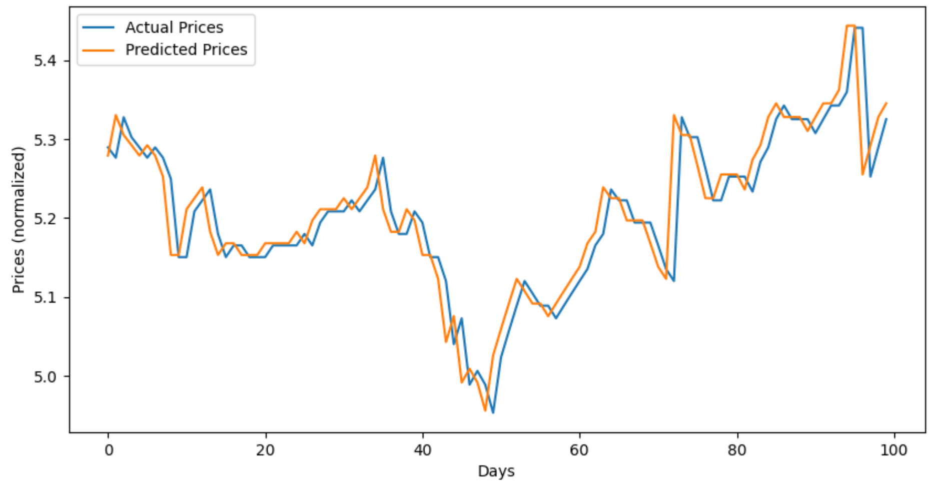

Figure 2.4: Actual vs Predicted Prices on AC for 100 days

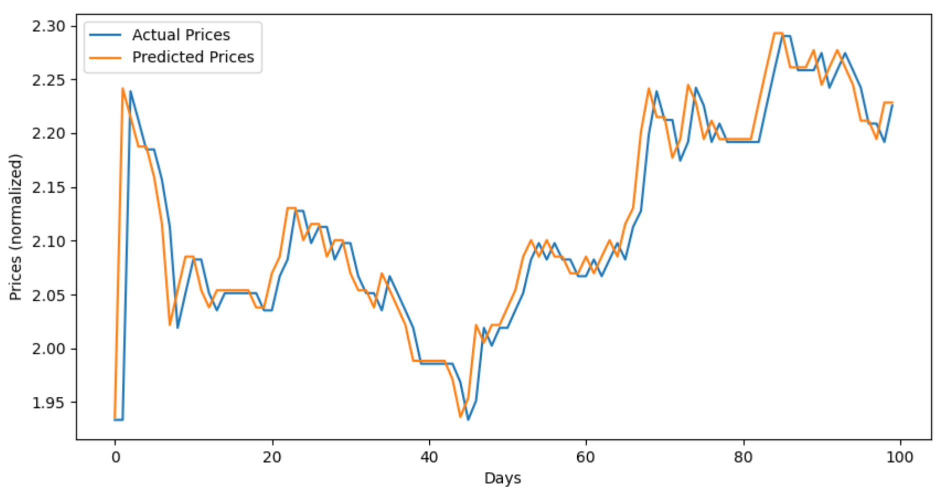

Figure 2.5: Actual vs Predicted Prices for ALI over 100 days

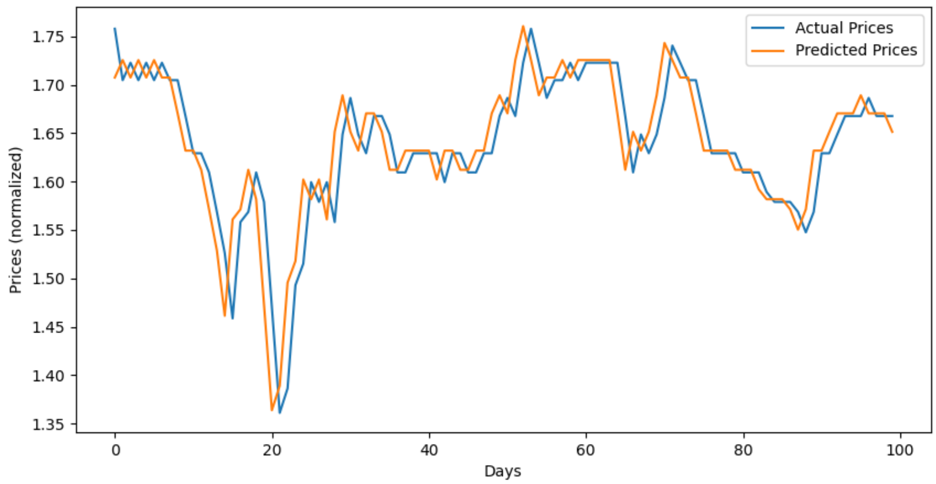

Figure 2.6: Actual vs Predicted Prices for AP over 100 days

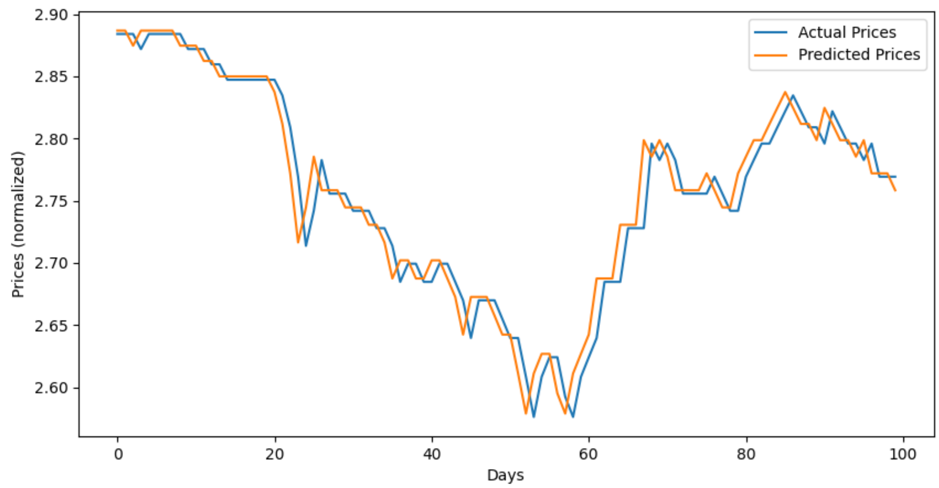

Figure 2.7: Actual vs Predicted Prices for BDO over 100 days

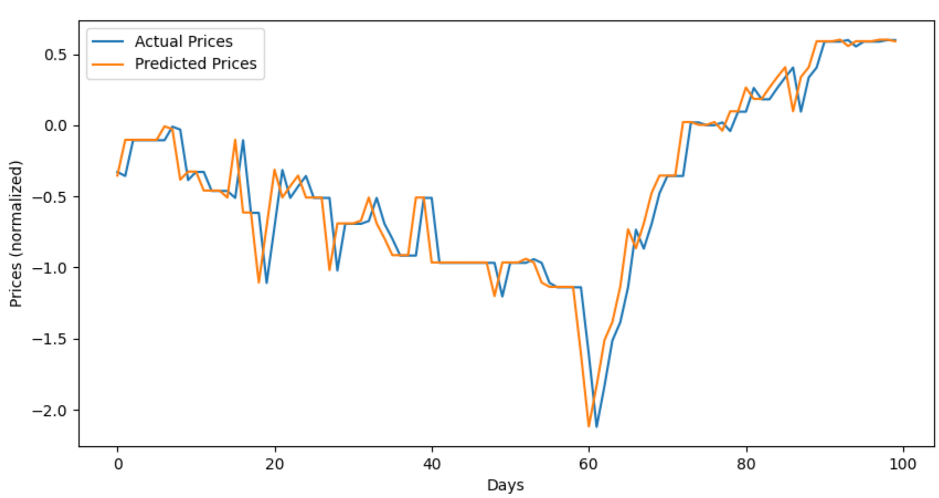

Figure 2.8: Actual vs Predicted Prices for BLOOM over 100 days

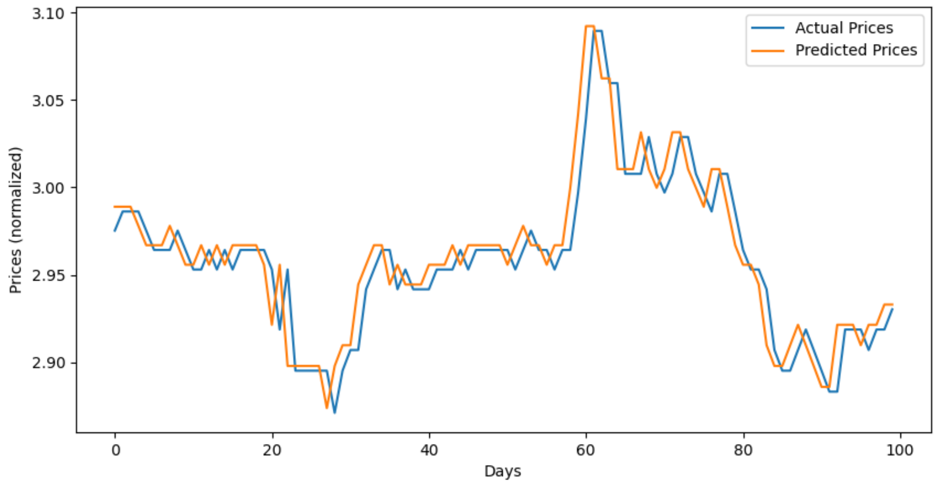

Figure 2.9: Actual vs Predicted Prices for FGEN over 100 days

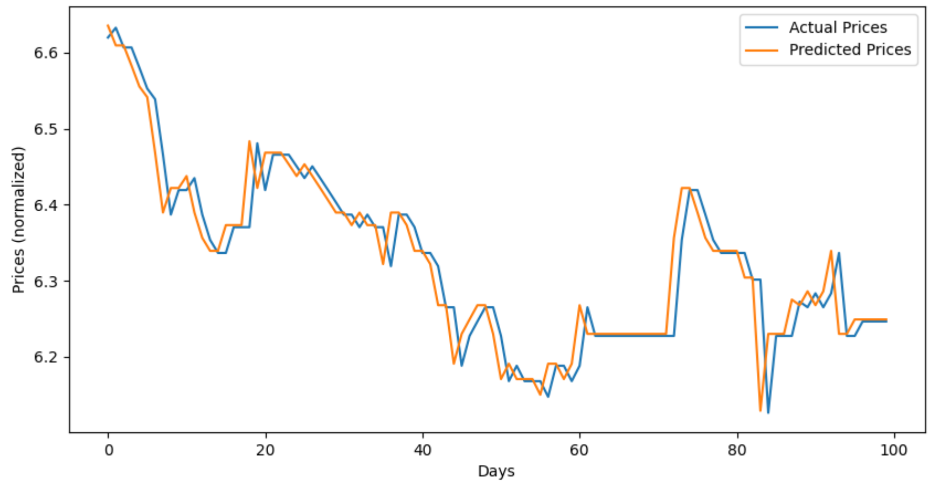

Figure 2.10: Actual vs Predicted Prices for GLO over 100 days

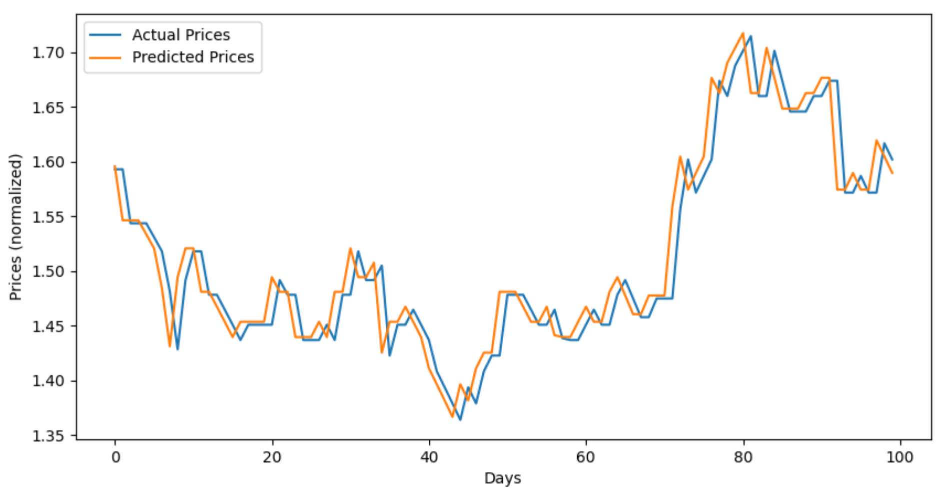

Figure 2.11: Actual vs Predicted Prices for ICT over 100 days

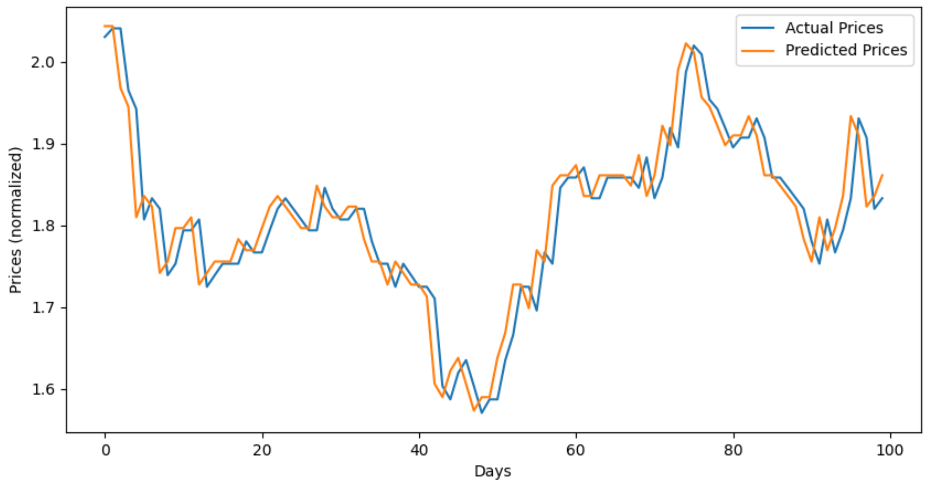

Figure 2.12: Actual vs Predicted Prices on JGS for 100 days

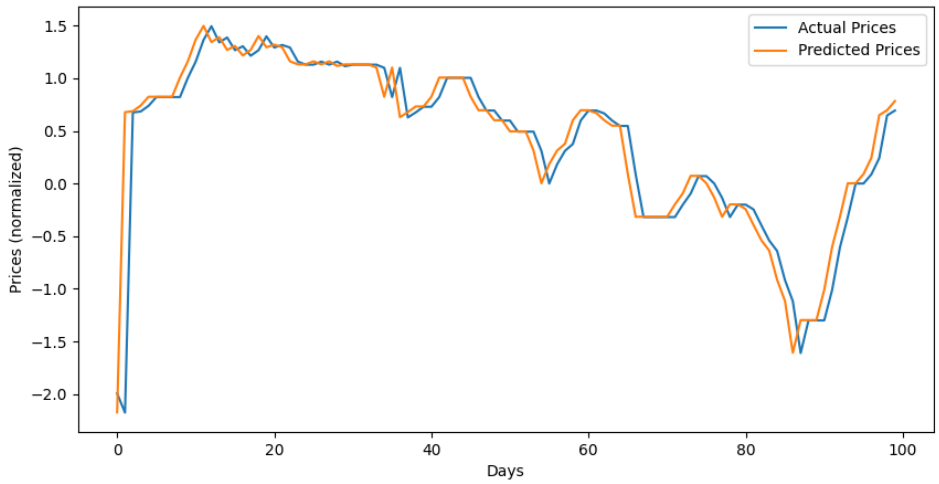

Figure 2.13: Actual vs Predicted Prices on LTG for 100 days

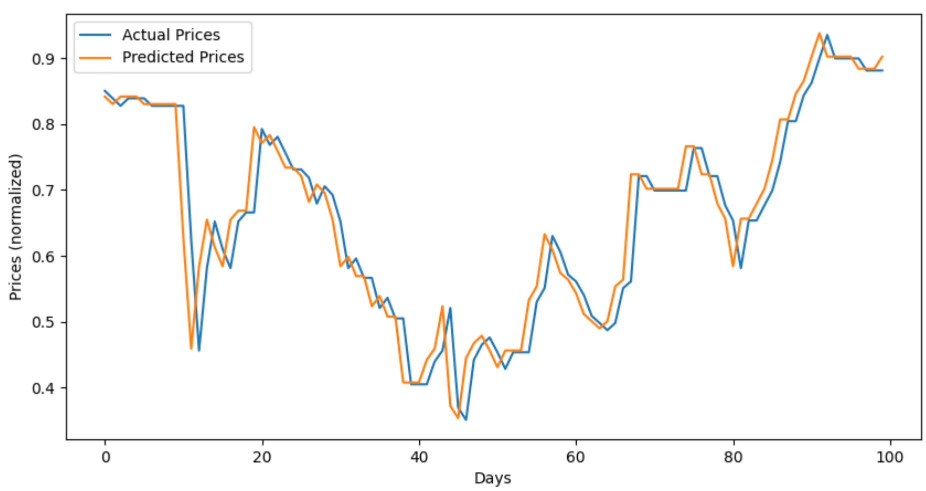

Figure 2.14: Actual vs Predicted Prices on MEG for 100 days

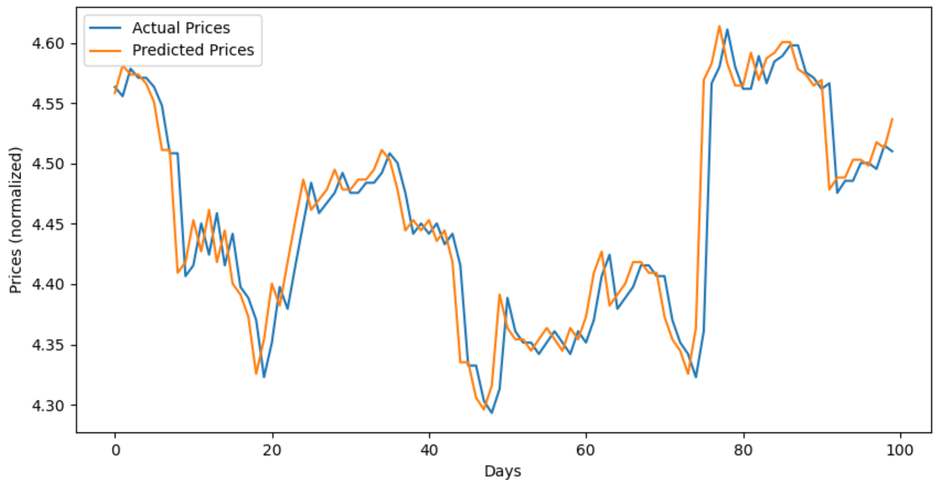

Figure 2.15: Actual vs Predicted Prices on MER for 100 days

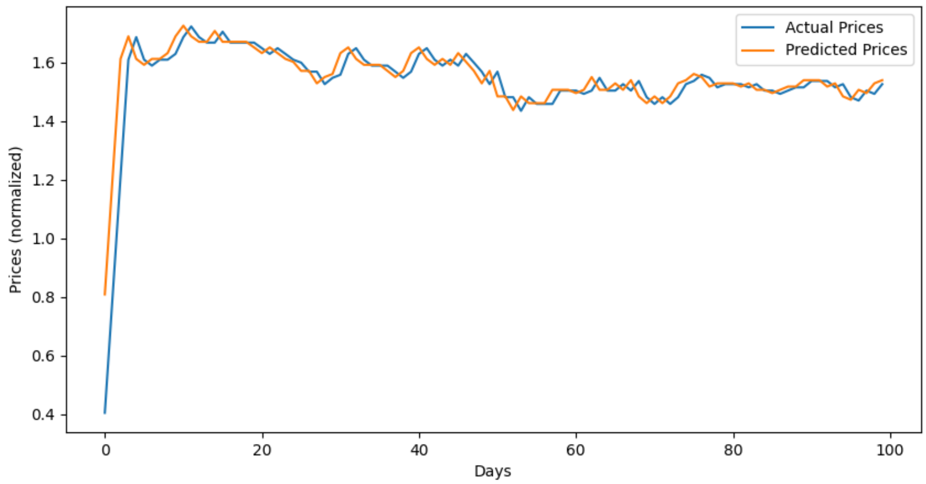

Figure 2.16: Actual vs Predicted Prices on MPI for 100 days

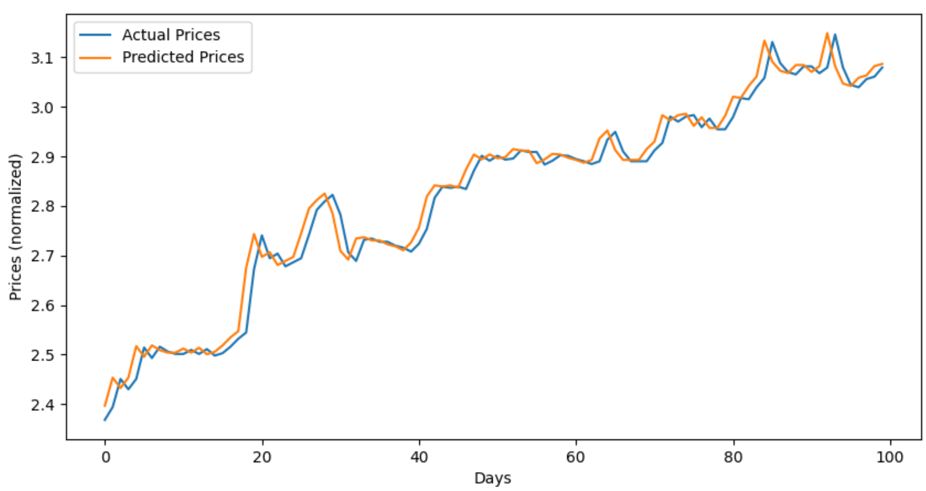

Figure 2.17: Actual vs Predicted Prices on PGOLD for 100 days

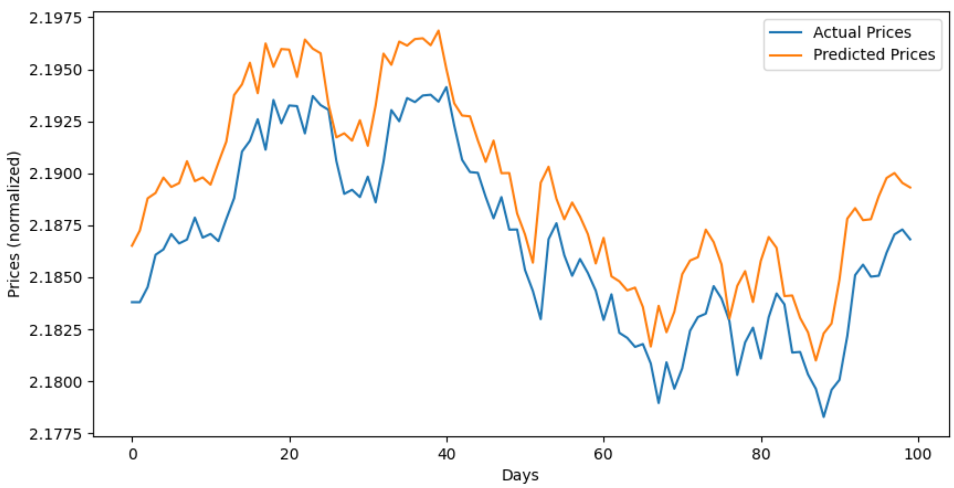

Figure 2.18: Actual vs Predicted Prices on PSEI for 100 days

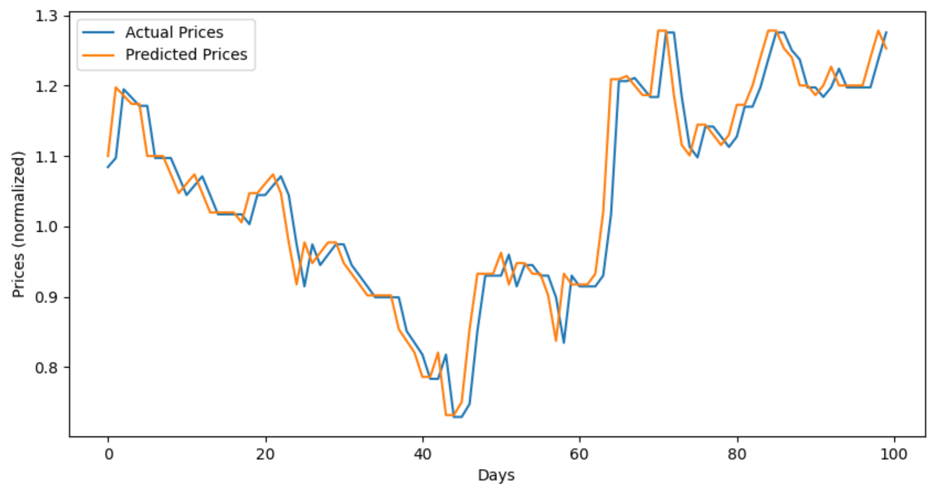

Figure 2.19: Actual vs Predicted Prices on RLC for 100 days

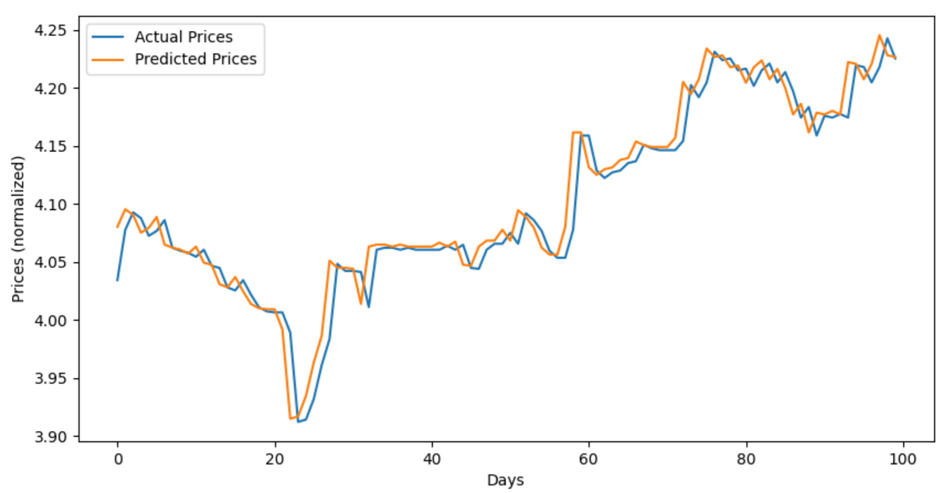

Figure 2.20: Actual vs Predicted Prices on RRHI for 100 days

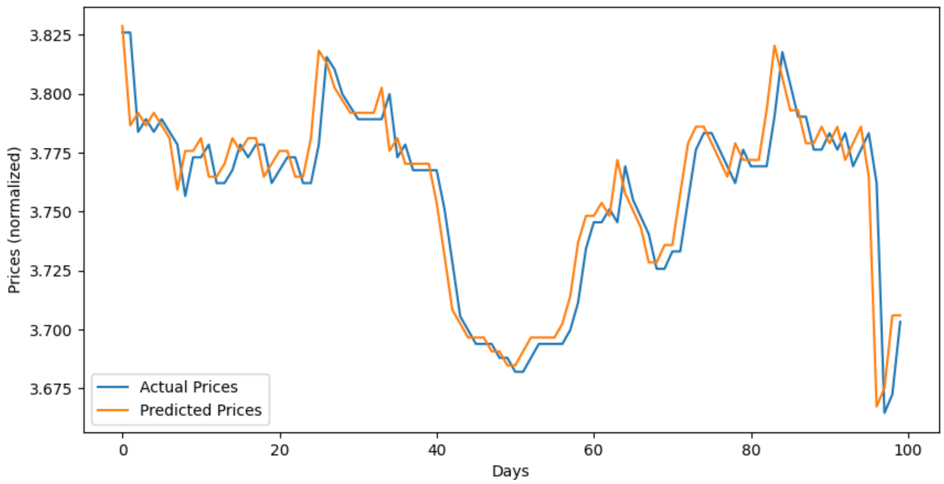

Figure 2.21: Actual vs Predicted Prices on SMC for 100 days

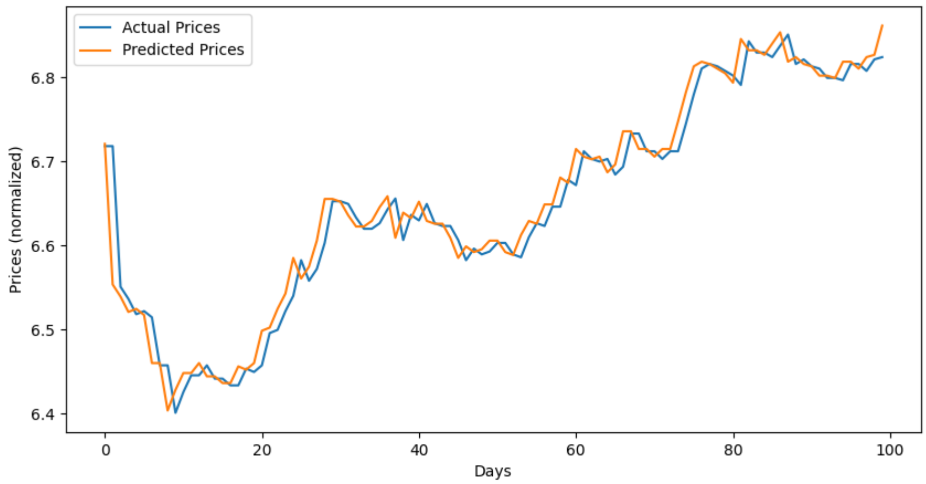

Figure 2.22: Actual vs Predicted Prices on TEL for 100 days

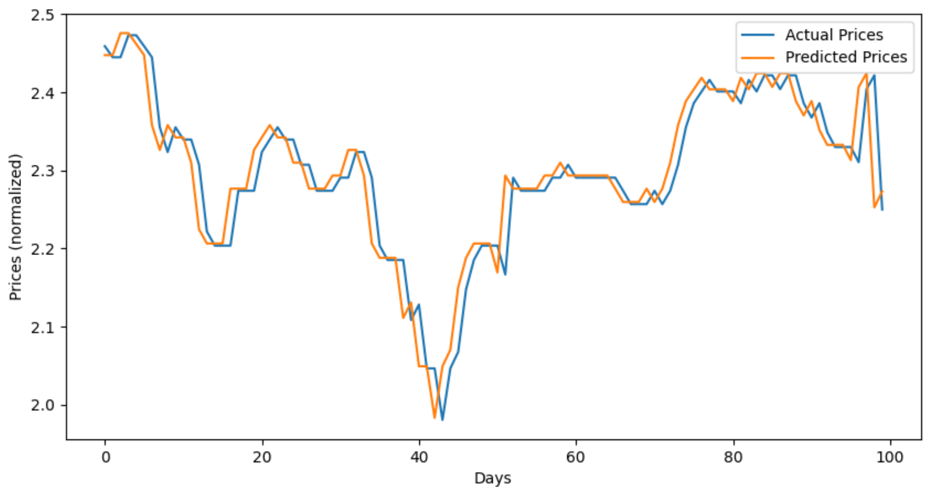

Figure 2.23: Actual vs Predicted Prices on URC for 100 days

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 2.8.

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 2.8, 2.11, 2.12, 2.13, and 2.14, 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.

Another observation from the graphs comparing actual and predicted prices over 100 days is that the predicted values appear to be higher than the actual prices. This indicates the possibility of loss because the model overestimates its prediction.

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 2.9 shows the MAPE scores for the successive predictions of the DMD-LSTM for each days.

Table 2.9: 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 it must be noted that the ratio values highlighted in red is to demonstrate that, despite the fact that negative ratio values shouldn't exist, doing so simply indicates that the data used to forecast the subsequent price data was overlapping by 2 to 5 times, depending on the ratio, and no longer used any actual data.

Moreover, in the integration of the DMD-LSTM model to the alamSYS, the 5 days successive predictions was utilized. Where it is shown from the Table 2.9 that it still performs well, even if the actual and predicted data ratio is only at 20%. This is also to limit the effect of stock market volatility that might affect the accuracy of the successive predictions of the model.

However, it can also be observed that the MAPE scores for successive days with zero to negative actual and predicted data ratio outperforms the MAPE scores from successive days 2 to 5 as illustrated in Figure 2.24, shown below.

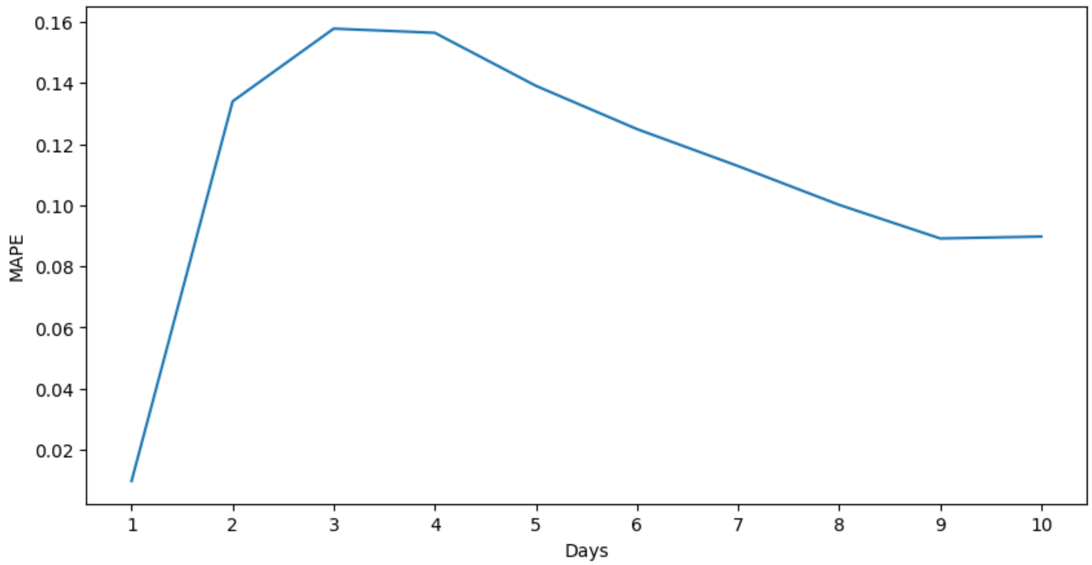

Figure 2.24: MAPE Scores for 1 to 10 (Days) Successive Predictions

Yet, since doing so might result in a poor generalization of data, they were not chosen to be the maximum consecutive days of predictions to be integrated in the alamSYS. As a matter of fact, it could be argued that these data's MAPE scores are overfitted, rendering them unreliable. On the contrary, it might also imply that the model maintains its accuracy for a longer time, even if the majority of the data used are those produced by the model itself. This could be a good thing, and may be attributed to the use of the dynamic modes, as first suggested in the study of Mann and Kutz (2015). In light of these considerations, additional testing is required to establish which of the two claims is true.

Overall, the results from the model training, evaluation, and cross-validation shows that the DMD-LSTM model developed in this special problem performs on par with the other studies that utilizes dynamic modes, as mentioned in Chapter **??** of this paper.

2.4 ALMACD Results and Discussions

The ability to predict consecutive days in the stock market is useless without a trading strategy - which allows risk mitigation and increases the probability of positive returns over time. Trading strategies, in particular, are based on a predefined set of rules and criteria that are used to determine when to buy and sell stocks. (Hayes, 2022).

A variety of algorithmic trading, on the other hand, refers to the use of mathematical and computational techniques to determine the best position to take for a specific set of stocks. Additionally, the possibility of loss due to the influence of human emotion is eliminated. (WallStreetMojo, n.d.).

Whereas, the author used the Arnaud Legoux Moving Average Convergence and Divergence (ALMACD) trading strategy in this special problem and integrated it into the alamSYS as the system's internal trading algorithm. ALAMCD uses predicted prices for the next 5 days, as well as 200 days of actual stock price data, to track the signals and output a simple flag indicating whether to buy or sell that stock at that time.

The compounded expected return after return backtesting is provided for each stock in Table 2.10 using the optimized parameters for the fast and slow ALMA. This was done to validate the potential returns for all stocks, not just the PSEI, from which the best ALMA parameters were derived.

Table 2.10: Optimal Alma Parameters Validation Results

Stock	Compounded Expected Return
PSEI	113966.8500
AC	20893.1914
ALI	1072.1418
AP	690.7100
BDO	2541.9970
BLOOM	495.4600
FGEN	581.0804
GLO	60538.0035
ICT	2815.6103

Table 2.10 continued from previous page

Stock	Compounded Expected Return
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

Based on the table of expected returns above, all stocks are expected to return a positive yield over time when these optimal ALMA parameters are used. It is also worth noting that the expected return is calculated for each unit of stock, which means that if we use the expected compounded return value of MPI at PHP 146.02, which appears to be the lowest - the actual return could be at least PHP 146,020, assuming the minimum board lot required for the stocks is 1000 shares (Pesobility, n.d.).

However, despite the high potential returns, investors should proceed with caution for two reasons. First, the expected return is based on historical price data, which may not follow the trend of future price data, potentially rendering the trading algorithm obsolete (Quantified Strategies, 2023). Second, the return calculation does not account for and compensate for the additional fees associated with buying and selling the stock, which can affect the overall actual returns. Moreover, the author investigated the potential for returns by following the alamSYS predictions, as discussed further in the succeeding section.

2.5 Results and Discussions for the Real World Application of alamSYS

xxx

Table 2.11: 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

References

Adams, R. (2022, 7). *Is Investing in Stocks Gambling? No, Investing Isn't Zero Sum.* Retrieved October 07, 2022, from https://youngandtheinvested.com/is-investing-in-the-stock-market-gambling/

Alegado, S., Lopez, D. B., & Calonzo, A. (2022, 8). *Philippine economic growth slows to 7.4% as inflation bites.* Retrieved October 06, 2022, from https://www.bloomberg.com/news/articles/2022-08-09/philippine-economy-expands-7-4-last-quarter-below-estimate

Bae, K. H., & Kang, J. (2017). *Does the stock market benefit the economy?* Retrieved from https://www.efmaefm.org/0EFMSYMPOSIUM/2017/papers/Does%20the%20Stock%20Market%20Benefit%20the%20Economy%20-%20updated.pdf

Balaba, J. M. (2017). *Does the stock market drive the philippine economy?* Retrieved from https://www.dlsu.edu.ph/wp-content/uploads/pdf/conferences/research-congress-proceedings/2017/RVREBM/RVREBM-I-003.pdf

Bangko Sentral ng Pilipinas. (n.d.). *Stock market.* Retrieved October 06, 2022, from https://www.bsp.gov.ph/Statistics/OtherRealSectorAccounts/stocks.pdf

Budiharto, W. (2021). Data science approach to stock prices forecasting in indonesia during covid-19 using long short-term memory (LSTM). *Journal of Big Data*, *8*(1). Retrieved from https://doi.org/10.1186%2Fs40537-021-00430-0 doi: 10.1186/s40537-021-00430-0

Bureau of the Treasury Bangko Sentral ng Pilipinas. (n.d.). *19 selected domestic interest rates weighted average in percent per annum for periods indicated.* Retrieved October 06, 2022, from www.bsp.gov.ph/Statistics/Financial%20System%20Accounts/tab19_dir.aspx

Business World. (2022, 5). *Psei sinks further as net foreign selling surges.* Retrieved October 06, 2022, from https://www.bworldonline.com/stock-market/2022/05/24/450559/psei-sinks-further-as-net-foreign-selling-surges/

Campbell, G. (2021, 5). *Does the stock market reflect the economy?* Retrieved October 07, 2022, from https://www.economicsobservatory.com/does-the-stock-market-reflect-the-economy

Canto, J., & Romano, K. (2022, 4). *Philippines economic outlook 2022.* Retrieved October 06, 2022, from https://www.mckinsey.com/featured-insights/asia-pacific/philippines-economic-outlook-2022

Chen, J. (2022, 7). *What is the stock market, what does it do, and how does it work?* Retrieved October 06, 2022, from https://www.investopedia.com/terms/s/stockmarket.asp

Commission on Population and Development. (2021). *Popcom: Number of filipinos in 2021 estimated at 110.8 million, sizes of families trending lower at 4 members.* Retrieved October 07, 2022, from https://popcom.gov.ph/popcom-number-of-filipinos-in-2021-estimated-at-110-8-million-sizes-of-families-trending-lower-at-4-members.

EdwardJones. (n.d.). *Why invest in stocks?* Retrieved September 19, 2022, from https://www.edwardjones.com/us-en/market-news-insights/guidance-perspective/benefits-investing-stock

Fayed, A. (2022, 4). *A list of 14 blue chip companies in the philippines.* Retrieved October 07, 2022, from https://adamfayed.com/a-list-of-14-blue-chip-companies-in-the-philippines/

Guo, Y. (2022). *Stock price prediction using machine learning.* Retrieved from https://www.diva-portal.org/smash/get/diva2:1672304/FULLTEXT01.pdf

Hall, M. (2022, 9). *How the stock market affects gdp.* Retrieved October 07, 2022, from https://www.investopedia.com/ask/answers/033015/how-does-stock-market-affect-gross-domestic-product-gdp.asp

Hayes, A. (2022). *What is a trading strategy? how to develop one.* Retrieved April 22, 2023, from https://www.investopedia.com/terms/t/trading-strategy.asp

Juviler, J. (2022). *Api response time, explained in 1000 words or less.* Retrieved April 22, 2023, from https://blog.hubspot.com/website/api-response

-time

Kim, T. (2022, 7). *Alternative data for hedge funds: The competitive edge to investing.* Retrieved November 04, 2022, from https://www.similarweb.com/corp/blog/investor/asset-research/hedge-funds-use-alternative-data

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022, 7). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, *197*, 116659. doi: 10.1016/j.eswa.2022.116659

Kuttichira, D. P., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017, 9). Stock price prediction using dynamic mode decomposition. In (p. 55-60). IEEE. doi: 10.1109/ICACCI.2017.8125816

Lu, H., & Tartakovsky, D. M. (2020, 1). Prediction accuracy of dynamic mode decomposition. *SIAM Journal on Scientific Computing*, *42*, A1639-A1662. doi: 10.1137/19M1259948

Mann, J., & Kutz, J. N. (2015, 8). Dynamic mode decomposition for financial trading strategies.

Pesobility. (n.d.). *Board lot of philippine stock exchange.* Retrieved April 22, 2023, from https://www.pesobility.com/ref/boardlot

Quantified Strategies. (2023). *Why trading strategies are not working (how to know when trading systems stopped performing?).* Retrieved April 22, 2023, from https://www.quantifiedstrategies.com/why-trading-strategies-are-not-working/

Rea, L. (2020, 8). *Predicting future stock market trends with python and machine learning.* Retrieved October 05, 2022, from https://towardsdatascience.com/predicting-future-stock-market-trends-with-python-machine-learning-2bf3f1633b3c

Reuters. (2022, 9). *Philippine economy to grow more slowly than previously thought in 2022, imf says.* Retrieved October 06, 2022, from https://www.reuters.com/markets/asia/philippines-economy-seen-growing-65-2022-50-2023-imf-2022-09-26/

Royal Bank of Canada Direct Investing Inc. (n.d.). Key benefits of investing in stocks. *Royal Bank of Canada Direct Investing Inc.*. Retrieved from www6.royalbank.com/en/di/hubs/investing-academy/chapter/key-benefits-of-investing-in-stocks/jv7atg13/jv7atg1j

Savaș, M. C. (2017). *Algorithmic trading strategies using dynamic mode decomposition: Applied to turkish stock market.* Retrieved from https://etd.lib.metu.edu.tr/upload/12621107/index.pdf

Schwab-Pomerantz, C. (2021, 2). *Is investing in the stock market gambling?* Retrieved October 07, 2022, from https://www.schwab.com/learn/story/is-investing-stock-market-gambling

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

Soni, P., Tewari, Y., & Krishnan, D. (2022, 1). Machine learning approaches in stock price prediction: A systematic review. *Journal of Physics: Conference Series*, *2161*, 012065. doi: 10.1088/1742-6596/2161/1/012065

Statista Research Department. (2022). *Number of stock market accounts philippines 2021.* Retrieved October 07, 2022, from www.statista.com/statistics/1194840/philippines-number-of-stock-market-accounts-by-type

Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y.-H. (2020). *Machine learning stock market prediction studies: Review and research directions* (Vol. 28). Retrieved from https://scholarworks.lib.csusb.edu/jitim/vol28/iss4/3

Summers, B. D. (2022, 6). *Investing vs. gambling understanding risk-adjusted.* Retrieved October 07, 2022, from https://www.forbes.com/sites/forbesfinancecouncil/2022/06/01/investing-vs-gambling-understanding-risk-adjusted-performance/?sh=1ca2a49957d9

The Economic Times. (n.d.). *What is 'stock market'.* Retrieved October 05, 2022, from https://economictimes.indiatimes.com/definition/stock-market

The Philippine Stock Exchange, Inc. (n.d.-a). *Company information.* Retrieved October 05, 2022, from https://edge.pse.com.ph/companyInformation/form.do?cmpy_id=478

The Philippine Stock Exchange, Inc. (n.d.-b). *Company list.* Retrieved October 05, 2022, from https://edge.pse.com.ph/companyDirectory/form.do

Trade Brains. (2022, 7). *How does the stock market affect the economy?* Retrieved October 07, 2022, from https://tradebrains.in/how-stock-market-affect-the-economy/

Trading Economics. (n.d.). *Philippines inflation rate.* Retrieved October 07, 2022, from https://tradingeconomics.com/philippines/inflation-cpi

U.S. Securities and Exchange Commission. (n.d.). *What are stocks?* Retrieved October 07, 2022, from www.investor.gov/introduction-investing/investing-basics/investment-products/stocks

WallStreetMojo. (n.d.). *Algorithmic trading.* Retrieved April 22, 2023, from https://www.wallstreetmojo.com/algorithmic-trading/

Zwetsloot, R. (2019). *Raspberry pi 4 specs and benchmarks.* Retrieved April 22, 2023, from https://magpi.raspberrypi.com/articles/raspberry-pi-4-specs-benchmarks