Available online at https://ganitjournal.bdmathsociety.org/

GANIT: Journal of Bangladesh Mathematical Society

GANIT *J. Bangladesh Math. Soc.* 44.1 (2024) 29–58
DOI: https://doi.org/10.3329/ganit.v44i1.73986

A Comparative Study of ARIMA, Artificial Neural Networks, and Kalman Filter Models for Dhaka Stock Exchange Forecasting

Anika Tahsin Biva *[a], Md Shafiul Alom Khan[a], and A B M Shahadat Hossain [a]

[a] *Department of Applied Mathematics, University of Dhaka, Dhaka-1000, Bangladesh*

ABSTRACT

The ARIMA, Artificial Neural Networks (ANN), Kalman Filter, and hybrid ARIMA ANN models are compared in this article in order to establish a forecast for some stock prices of the Dhaka Stock Exchange (DSE). From December 2021 to November 2023, the closing prices of ten companies listed on the DSE comprised the dataset for this study. We employ root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) to determine which model is better suited and to assess the performance of the models. The main aim of our study is to show how these models compare for these ten companies and to determine which one is better for each of them. This is because the datasets used in this study are diverse; some of the companies' datasets have linear trends or seasonal patterns, some have nonlinear patterns, and some have both trend and complex nonlinear interactions. In this study, we use **Python** and **R** to compare our obtained results both numerically and graphically.

Received: January 24, 2024 **Accepted:** May 21, 2024 **Published Online:** June 30, 2024

Keywords: ARIMA; ANN; KF; RMSE; MAE; MAPE
AMS Subject Classifications 2024: 62M10, 68T05, 62M20.

Nomenclature	
ARIMA	AutoRegressive Integrated Moving Average
KF	Kalman Filter
AAMRATECH	Aamra Technologies Limited
ADVENT	Advent Pharma Limited
ARGONDENIM	Argon Denims Limited
BDTHAIFOOD	BD Thai Food & Beverage
ECABLES	Eastern Cables Ltd.
KEYACOSMAT	Keya Cosmetics Ltd
MEGHNAPET	Meghna Pet Industries
PRIMETEX	Prime Textile Spinning Mills Limited
UNIONCAP	Union Capital Limited
PUBALIBANK	Pubali Bank Ltd.

*Corresponding author. *E-mail address:anika19992017@gmail.com*

1 Introduction

As Bangladesh's main stock exchange, DSE acts as a market for securities trading, offering a venue for the purchase and sale of stocks, bonds, mutual funds, and other financial products. The final traded price of a specific stock or security at the conclusion of the trading day is referred to as the closing price on the DSE [21]. Because it is made up of consecutive observations that are recorded at regular intervals typically daily it is regarded as time series data. Sequential data points are gathered at regular intervals to form time series, and time series forecasting is essential for projecting future trends and supporting risk management, resource planning, and decision-making in a variety of fields [10],[15].

Forecasting entails analyzing past data to make predictions about future events. Investors who engage in active trading in the stock market rely heavily on anticipating the price of particular stocks and market behavior by analyzing past price movements. This analysis greatly influences their decision making process. An investor's potential benefit increases in direct proportion to the accuracy of a projection. Conversely, if the forecast is incorrect, it might potentially lead to a substantial financial loss. Stock market forecasting is widely regarded as a challenging task in the field of finance due to the unpredictable nature and intricate relationships within the stock market [9].

Multiple research projects have been undertaken in the financial domain to predict the future path of stock prices. Various studies have determined that it is possible to accurately forecast stock values by employing well-designed statistical, econometric, and machine learning models [35]. Sen and Datta almost accurately predicted the stock price using time series decomposition [19]-[20], [24] -[27]. Again many researchers like Hossan et al. have studied the numerically to understand the impact of European options with transaction costs [42]. Al Mobin and Kamrujjaman conducted an analysis focused on employing algorithms to improve forecast accuracy through the downsizing of epidemiological time series data [43]. Other researchers have conducted a comparison of results using predictable and stochastic analysis methods [44]. We have studied their findings for this study to ensure its significance and credibility.

For capturing linear relationships and patterns in stock market data, ARIMA models work well. In order to predict future prices, they examine past price movements, looking for trends, seasonality, and autocorrelations. When the data shows stationary behaviour, ARIMA models work well and can shed light on short to medium term price changes [2].

ANNs are effective at capturing complex and nonlinear interactions in stock market data. They can detect hidden trends that typical linear models may overlook. Large datasets can be handled by ANNs, which can also be used to adapt to shifting market conditions and understand complex correlations between different market indicators and stock prices, all of which have the potential to increase forecasting accuracy [12].

When working with imperfect or noisy data, the Kalman Filter is employed to determine the system's state. In order to make more accurate predictions of underlying trends or market conditions, the Kalman Filter may be used to stock market data in order to reduce noise. Particularly useful when dealing with data that is updated in real-time or when predictions need to be revised often to account for new information [8].

The hybrid ARIMA ANN model combines ANN's capacity to model nonlinear patterns with ARIMA's strengths in capturing linear relationships. By utilising both approaches, this hybrid strategy may be able to provide prediction accuracy improvements. For example, whereas ANN can strengthen forecasts by capturing more complicated linkages and producing a more thorough forecast, ARIMA can capture initial trends or patterns [14],[16],[27].

Our study provides a comprehensive comparison of multiple forecasting models, including ARIMA, ANN, Kalman Filter, and a hybrid ARIMA ANN model, specifically for the DSE. Unlike previous research that often focuses on a single model or a limited dataset, our work analyzes the performance of these diverse models using closing prices from ten DSE-listed companies over a two-year period. By employing rigorous evaluation metrics such as RMSE, MAE, and MAPE, and utilizing both Python and R for analysis, we offer a detailed assessment of each model's suitability for different types of data patterns. This study not only highlights the strengths and weaknesses of each forecasting method but also provides practical insights for investors and analysts in making more informed decisions, thereby advancing the field of financial forecasting.

The main objectives of our study are:

- We compare the performance of ARIMA, ANN, Kalman Filter, and hybrid ARIMA ANN models in forecasting stock prices for companies listed on the DSE.
- We analyze the closing prices of ten DSE-listed companies, covering the period from December 2021 to November 2023.
- We evaluate each forecasting model's performance using RMSE, MAE, and MAPE.
- We determine the most suitable forecasting model for each company, considering the diversity in their datasets, which may include linear trends, seasonal patterns, nonlinear patterns, and complex interactions.

Our study is designed as follows: Section 2 outlines the methodology used in the study. It details the data collection procedure, explaining how the closing prices of ten companies listed on the DSE were gathered from December 2021 to November 2023. It provides a brief description of the forecasting models used: ARIMA, ANN, Kalman Filter, and the hybrid ARIMA ANN method. It explains how these methods are applied to forecast the closing prices of the selected companies. The section also describes the evaluation metrics used to assess model performance: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Section 3 presents the results of the forecasting models. It includes a detailed analysis of the performance of each model for the ten DSE companies. The results are compared using the evaluation metrics described in Section 2. It provides both numerical and graphical analyses to illustrate the accuracy and effectiveness of each forecasting method. The section discusses the implications of the findings, highlighting the strengths and weaknesses of each model in the context of stock price prediction. Section 4 summarizes the key findings of the study. It highlights which forecasting models performed best for different types of stock price data. The conclusion discusses the potential applications of the research findings. It suggests directions for future work, emphasizing the importance of combining different forecasting models to improve prediction accuracy in financial markets.

2 Methodology

2.1 Data Collection

This paper's dataset was collected from the DSE website [30]. We gathered comprehensive stock market information for each of the 392 DSE companies with **Python** coding. The closing prices of 10 DSE companies are taken into consideration in this study. AAMRATECH , ADVENT, ARGONDENIM, BDTHAIFOOD, ECABLES, KEYACOSMET, MEGHNAPET, PRIMETEX, UNIONCAP, PUBALIBANK are the companies in concern. Closing prices from December 26, 2021 to September 30, 2023 are used as training data, while closing prices from October 1, 2023 to November 30, 2023 are used as test data. Out of a total of 474 observations, 430 are utilised for training purposes and 44 for testing purposes.

2.2 ARIMA

ARIMA stands for autoregressive integrated moving average. It is a well-liked statistical technique for forecasting and analysing time series. ARIMA models are capable of capturing many different temporal structures and patterns in data.

Autoregressive, integrated, and moving average are the three components that make up ARIMA. An autoregressive (AR) component plots the series' current values against its past ones. The present value of the series is assumed to be a linear mixing of its earlier values [36]. The goal of the integrated (I) component is to achieve stationarity in the series by using differencing, which ensures a consistent mean and variance throughout time. Stationarity is critical for many time series models, such as ARIMA, as they depend on the assumption that the statistical properties of the data remain constant throughout time. The "I" in ARIMA represents the number of variations needed to achieve stationarity. The association between an observation and the residual error in a moving average model, utilised for examining lagged data, is shown by a moving average (MA) component [1] [3].

An ARIMA model is represented by the notation ARIMA(p, d, q), where ‘p’ denotes the autoregressive order, ‘d’ represents the degree of differencing (indicating the number of times the data has been adjusted to achieve stationarity), and ‘q’ signifies the moving average order [4] [5] [40].

We can write ARIMA(p, d, q) model is as follows:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2.1}$$

where:

- y_t is the value of the time series at time ‘t’.
- c is the constant term.
- $\phi_1, \phi_2, ..., \phi_p$ are the parameters of the autoregressive part.
- $\theta_1, \theta_2, ..., \theta_q$ are the parameters of the moving average part.
- ε_t is the error term at time ‘t’.

Based on patterns seen in the data, the ARIMA model forecasts future values using historical data. The parameters ϕ and θ need to be estimated using methods like least squares and maximum likelihood estimation.

ARIMA models use moving average, autoregression, and differencing to find patterns and dependencies in time series data for analysis and forecasting. The Augmented Dickey-Fuller (ADF) test and the Durbin Watson (DW) test are critical tools for fitting and testing these models [2].

In our study, for forecasting with ARIMA we first perform an ADF test to check whether the data is stationary or not, because ARIMA is only applicable to stationary datasets. We consider the confidence level is set to 95%, indicating that the forecasted values will be accompanied by prediction intervals representing the uncertainty in the forecast. This means that the intervals will cover the range within which future observations are likely to fall with a 95% probability.

2.3 Artificial Neural Network

The artificial neural network (ANN) is a computer model based on the architecture and functioning of neural networks as seen in the human brain. It’s a machine learning method that can identify intricate relationships and patterns in data, which makes it useful for a variety of tasks like pattern recognition, regression analysis, and classification [13].

An ANN comprises interconnected nodes known as neurons organized in layers:

1. **Input Layer**: Neurons in this layer receive input data features.
2. **Hidden Layers**: These intermediate layers process information by applying weights to inputs and passing them through activation functions.
3. **Output Layer**: Neurons in this layer produce the network’s output based on the processed information.

Each connection between neurons has an associated weight that gets adjusted during the learning process, allowing the network to learn from the data.

Let’s say we have ‘n’ input features represented by $x_1, x_2, ..., x_n$.

Associated with each input feature, there are weights denoted by $w_1, w_2, ..., w_n$.

Additionally, there’s a bias term denoted by ‘b’.

The weighted sum of inputs plus the bias is computed:

$$z = w_1x_1 + w_2x_2 + ... + w_nx_n + b \tag{2.2}$$

Next, an activation function (often denoted as ϕ) is applied to this weighted sum 'z' to introduce non-linearity into the neuron's output:

$$a = \phi(z) \tag{2.3}$$

The output 'a' is the result after passing through the activation function. Common activation functions include:

- **Sigmoid**: $\sigma(z) = \dfrac{1}{1+e^{-z}}$
- **ReLU (Rectified Linear Unit)**: $\text{ReLU}(z) = \max(0, z)$
- **Tanh**: $\tanh(z) = \dfrac{e^z - e^{-z}}{e^z + e^{-z}}$

For a neural network with multiple layers, the output of each neuron in a layer becomes the input to each neuron in the subsequent layer. This process is repeated until reaching the output layer.

Given an input x, the output of a neuron in the first hidden layer can be expressed mathematically as:

$$a^{[1]} = \phi(W^{[1]}x + b^{[1]}) \tag{2.4}$$

where:

- $W^{[1]}$ represents the weights of the connections between the input layer and the first hidden layer.
- $b^{[1]}$ is the bias term for the first hidden layer.
- $a^{[1]}$ is the output of the first hidden layer.

This process persists via successive concealed levels until the ultimate output layer is attained, which generates the network's output [17].

During the learning process, neural networks use algorithms like backpropagation (adjusting weights based on prediction errors) and optimization techniques such as gradient descent to minimize a predefined loss or error function. The model iteratively adjusts the weights and biases to improve its predictions on the training data [23].

Artificial neural networks (ANNs) are utilized for time series forecasting due to their ability to capture complex nonlinear relationships present in the data [18]. Time series data frequently include complex connections and patterns that are difficult to identify using conventional statistical techniques. From historical time series data, ANNs are excellent at learning and recognizing these complex patterns because of their numerous layers and nonlinear activation functions. They are able to forecast future time steps by identifying seasonality, dependencies, and hidden trends in the data. They can also handle enormous volumes of data effectively, adapt to changing patterns, and generalize well to new unseen data [41].

In our study

- A sequential model from the Keras API (Application Programming Interface) is used to construct ANNs. This model allows us to build linear stacks of layers, similar to feedforward neural networks, where layers are placed one after another.
- Since this is a univariate time series dataset with a singular feature (closing prices) that we are utilizing to forecast stock prices, the input layer consists of just one neuron.
- The network consists of two concealed, substantial levels. The first hidden layer consists of 64 neurons that employ the Rectified Linear Unit (ReLU) activation function, while the second hidden layer has 32 neurons that also utilize the same activation function. ReLU activation is selected for hidden layers due to its effectiveness in handling non-linearities and preventing the vanishing gradient problem [37].

- The output layer consists of a single neuron, suitable for regression tasks where predicting a continuous value (closing price) is the objective. No explicit activation function is specified, which implies a linear activation appropriate for regression tasks to output unbounded values.
- The model compilation loss function is defined as the 'mean squared error', which quantifies the average of the squared discrepancies between the predicted and actual values. The Mean Squared Error (MSE) is a commonly utilized metric for regression tasks due to its ability to assign greater weight to bigger mistakes, hence making it well-suited for forecasting stock values.
- The Adam optimizer adapts the learning rate during training, resulting in faster convergence and better handling of sparse gradients.
- Early stopping is employed as a measure to mitigate overfitting by terminating the training process if there is no discernible improvement in the validation loss. This ensures that the model is capable of generalizing effectively to unfamiliar data.

2.4 Kalman Filter

The Kalman Filter is a mathematical method used to estimate and predict the state of a system as it changes over time. Fundamentally, the Kalman Filter operates by continuously incorporating new measurements (which might contain noise or inaccuracies) to update its estimation of the current state of a system. Based on an incoming data set and a dynamic model of the system, it iteratively improves its predictions using a series of recursive equations [8],[38].

The filter operates in two stages: prediction and updating.

1. **Prediction**: To guess what the system will do next, the Kalman Filter looks at its behavior and previous states, which are generally shown by a state transition matrix and control inputs.
2. **Update**: Once new measurements become available, the filter compares these measurements with the predicted state. It then combines the predicted state with the measurement using a weighted average, giving more weight to measurements with higher certainty (lower noise).

The Kalman filter is a useful tool for time series predicting because it lets us guess what values will be in the future by using past data and quickly adding new data. By iteratively revising its estimations using both past patterns and present observations, it offers more precise predictions in comparison to some conventional approaches. The algorithm is adept at managing incomplete data, inconsistent time intervals between samples, and noisy measurements, rendering it appropriate for practical scenarios where the data may not be entirely dependable [39].

Mathematically, the Kalman Filter is a recursive algorithm for finding the best estimate of $\utilde{x}_t$ from $\utilde{y}_1$,$\utilde{y}_2$,$\cdots$,$\utilde{y}_t$, when the signal $\{\utilde{x}_t\}$ satisfy

$$\utilde{x}_t = F\utilde{x}_{t-1} + \utilde{v}_t \text{ (state equation)}$$

and the observations $\{\utilde{y}_t\}$ satisfy

$$\utilde{y}_t = G\utilde{x}_t + \utilde{w}_t \text{ (observation equation)}$$

It is assumed that

(a) $\{\utilde{v}_t\}$ are i.i.d multivariate normal with mean vector $\utilde{0}$ and covariance matrix Q;

(b) $\{\utilde{w}_t\}$ are i.i.d multivariate normal with mean vector $\utilde{0}$ and covariance matrix R;

(c) $\{\utilde{w}_t\}$ are independent of $\{\utilde{v}_t\}$

Then $\utilde{x}_t$ is a linear function of $\utilde{v}_t$, $\utilde{v}_{t-1}$,$\cdots$ and $\utilde{y}_t$ is a linear function of $\utilde{w}_t$,$\utilde{v}_t$ $\utilde{v}_{t-1}$,$\cdots$

Therefore $\utilde{x}_s$, $\utilde{y}_s \perp \utilde{v}_t$, $\utilde{w}_t$ when $s < t$, and $\utilde{x}_s \perp \utilde{w}_t$ any s, t

2.5 Hybrid ARIMA ANN

Linear and nonlinear components can both be present in time series data. For time series data, ARIMA is efficient at capturing linear trends and seasonality, whereas neural networks are better at capturing complex non-linear patterns [13]. Combining nonlinear and linear models into a hybrid approach improves forecast accuracy for time series data compared to standalone models [29].

Several hybrid methodologies outlined in the literature incorporate the subsequent approach: The ARIMA model is promptly utilized for the provided time series data. The residuals obtained from the ARIMA model are considered to be a non-linear component, and this non-linear data is then modeled using artificial neural networks (ANN) employing various methodologies.

Zhang [23] introduced a hybrid model that implies the time-series y_t can be represented as the sum of both linear and nonlinear components:

$$y_t = L_t + N_t \tag{2.5}$$

The ARIMA model is utilized to estimate the linear component L_t of y_t, and the predictions $\hat{L}_t$ are derived accordingly:

$$\hat{L}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \tag{2.6}$$

The above equation involves subtracting the linear forecasts generated by the ARIMA model from the initial series y_t.

Zhang [23] emphasised that the residuals series should consist of nonlinear variations, while the ARIMA model is only suitable for accurately fitting linear variations.

$$N_t = y_t - \hat{L}_t \tag{2.7}$$

The residual series N_t is modeled and forecasted using artificial neural networks (ANN), while the predictions $\hat{N}_t$ are derived using the following equation:

$$\hat{N}_t = f\left(n_t, n_{t-1}, \ldots, n_{t-n}\right) + v_t \tag{2.8}$$

The variable $\hat{N}_t$ represents the predicted nonlinear error-series. The function f is a nonlinear function that depends on the previous residuals. The variable v_t represents the component white-noise in the artificial neural network (ANN) modeling.

The ultimate hybrid model predictions are derived by adding the ARIMA forecasts with the ANN predictions, as shown in the following equation:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{2.9}$$

2.6 Model Evaluation

Throughout the evaluation phase, determining the best prediction model with the highest level of accuracy is crucial. RMSE , MAE, and MAPE are often used measures to assess forecasting model accuracy. They aid in assessing the disparities between expected and actual values. These measures with lower values suggest higher prediction performance. These measures aid in the selection of the best-performing model among alternatives by taking into account how close their forecasts are to the actual values. Rather than depending on just one statistic, it is critical to employ numerous metrics in tandem to acquire a thorough knowledge of a model's performance [14].

1. Root Mean Square Error (RMSE): Root Mean Square Error (RMSE) is a widely employed statistic that quantifies the average magnitude of the discrepancies between expected and actual values. The calculation involves finding the square root of the average of the squared discrepancies between the expected and actual values.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{2.10}$$

Here, Y_i represents the actual values, $\hat{Y}_i$ represents the predicted values, and n is the number of observations.

RMSE quantifies the accuracy of the model, with lower RMSE values indicating superior prediction ability. The presence of the squared word makes it highly susceptible to significant mistakes.

2. Mean Absolute Error (MAE): Mean Absolute Error (MAE) computes the mean of the absolute disparities between anticipated and actual values. It provides a more straightforward measure of error without considering the direction (overestimation or underestimation).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{2.11}$$

Similar to RMSE, lower MAE values indicate better accuracy. MAE is not as sensitive to outliers as RMSE due to the lack of squaring the differences.

3. Mean Absolute Percentage Error (MAPE): MAPE quantifies the percentage deviation between projected and actual values, offering an error percentage in relation to the actual values. It is especially beneficial when we want comprehension of the precision in percentage terms.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100 \tag{2.12}$$

Lower MAPE values indicate better accuracy. However, MAPE can pose challenges when dealing with actual values close to zero since division by zero results in undefined or infinite values.

3 Results and Discussions

Within this part, we conduct extrapolatory data analysis on a set of 10 DSE companies. The descriptive statistics for both the training and test datasets are displayed. We generate visual representations of the training and test datasets. We conduct an Augmented Dickey-Fuller (ADF) test to determine the stationarity of the training dataset. If the dataset exhibits stationarity, we can determine the appropriate ARIMA (p,d,q) model. We do a Durbin Watson test to determine the presence of autocorrelation in the training dataset. We visually depict both the observed and predicted values for the test dataset using the ARIMA, ANN, KF, and hybrid ARIMA ANN models. In order to assess the performance of the models, we compute the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) for each model.

3.1 AAMRATECH

Aamra Technologies Ltd. provides information technology solutions. The company launched on March 14, 1990, and its headquarters are in Dhaka, Bangladesh. It was enlisted in the DSE on July 4, 2012 [32].

Table 3.1: Statistics for Training Dataset of AAMRATECH

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	430	27.3	56.9	36.30790697674419	35.6	4.343779711549458

Table 3.2: Statistics for Test Dataset of AAMRATECH

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	44	30.4	32.1	30.99090909090909	30.9	0.5006127323814392

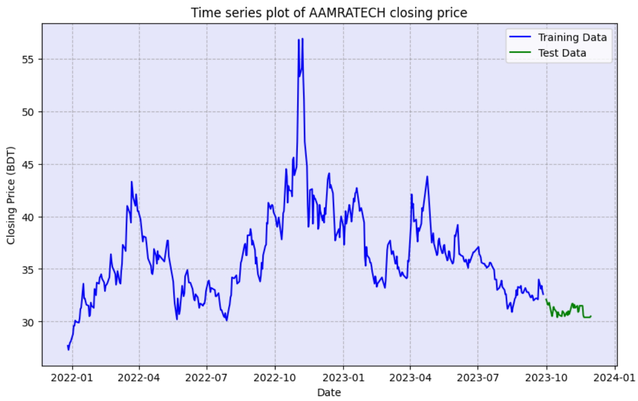


Figure 3.1: Training and Test Dataset of AAMRATECH

Table 3.3: ADF test, DW test, ARIMA model selection result for AAMRATECH

Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.1464 (stationary)	1.897263(no autocorrelation)	ARIMA(0,1,0)

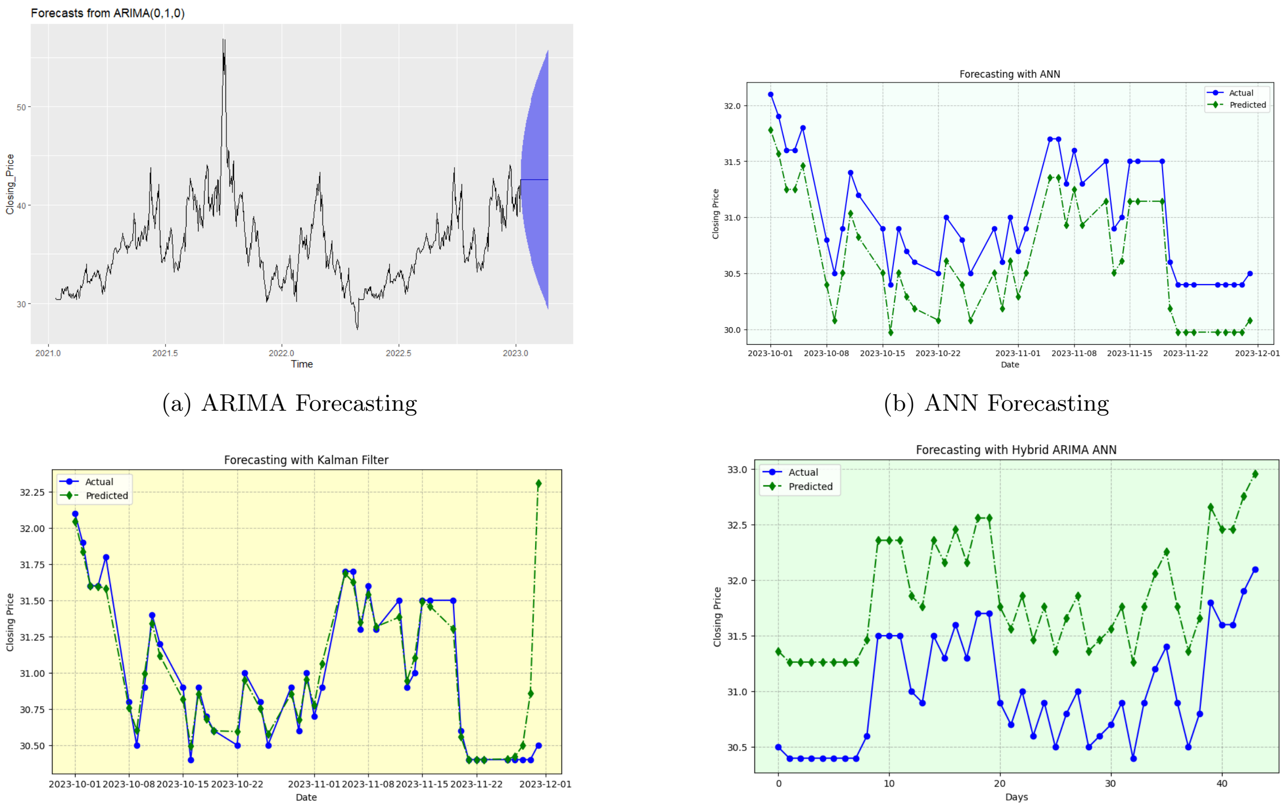
Table 3.4: Model Performance for AAMRATECH

Model	RMSE	MAE	MAPE
ARIMA	11.619634636722	11.6090909090909	0.374944913038945
ANN	0.38858809316855264	0.3874231598593969	0.012519746684119155
Kalman Filter	0.2916	0.1091	0.0035
HYBRID ARIMA ANN	0.3262989146336449	0.32560756856745016	0.010520081473575712

According to the data in table 3.3, we can see that the training dataset is stationary. The training datasets exhibit no autocorrelation. By examining tables 3.1 and 3.2, it is evident that the standard deviation of the training dataset surpasses the standard deviation of the test dataset. If the standard deviation of the training data is greater than the standard deviation of the test data, it indicates a discrepancy in the variability or spread of the two datasets, which can be attributed to random fluctuations. Figure 3.1 demonstrates that the training data displays a mixture of a persistent linear pattern and temporary nonlinear variations. Figure 3.2 visually illustrates the results of the ARIMA, ANN, KF, and hybrid ARIMA ANN models. Based on the data shown in table 3.4, it is evident that the Kalman Filter exhibits superior accuracy compared to the other three methods, making it the most optimum choice for predicting. Therefore, we compare the accuracy of predicting future outcomes using hybrid ARIMA ANN, ANN , and ARIMA . While the hybrid ARIMA ANN model has a slightly lower error rate compared to the ANN model, both hybrid ARIMA ANN and ANN, provide better performance for AAMRATECH. ARIMA is not suitable for AAMRATECH due to its significant level of inaccuracy.

3.2 ADVENT

Advent Pharma Ltd. is a pharmaceutical product manufacturer, importer, and marketer. It provides nutritional supplements, medications for animal health, and livestock feed additives in powder, bolus, and liquid dosage forms. The company's headquarters are in Dhaka, Bangladesh, and it was established on January 25, 2007. It was enlisted in the DSE on April 12,2018 [32].



(a) ARIMA Forecasting

(b) ANN Forecasting

(c) KF Forecasting

(d) Hybrid ARIMA ANN Forecasting

Figure 3.2: Actual vs Predicted Closing Price of AAMRATECH

Table 3.5: Statistics for Training Dataset of ADVENT

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	430	22.0	31.0	25.443023255813955	25.2	2.099835752133507

Table 3.6: Statistics for Test Dataset of ADVENT

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	44	22.6	23.9	22.765909090909094	22.7	0.2514545423971843

Table 3.7: ADF test, DW test, ARIMA model selection result for ADVENT

Augmented Dickey Fuller test (p value)	**Durbin Watson test Statistics**	**ARIMA (p,d,q)**
0.01 (stationary)	2.020704(no autocorrelation)	ARIMA(3,0,2)

From table 3.7, we observe that the training dataset is stationary. The training datasets exhibit no autocorrelation. Tables 3.5 and 3.6 indicate that the standard deviation of the training dataset is greater than the standard deviation of the test dataset, which can be attributed to random fluctuations. Based on the analysis of figure 3.3, it is evident that the training data displays recurring patterns or movements over a period of time. Figure 3.4 graphically represents the outcomes of ARIMA, ANN, KF, and hybrid ARIMA ANN. From

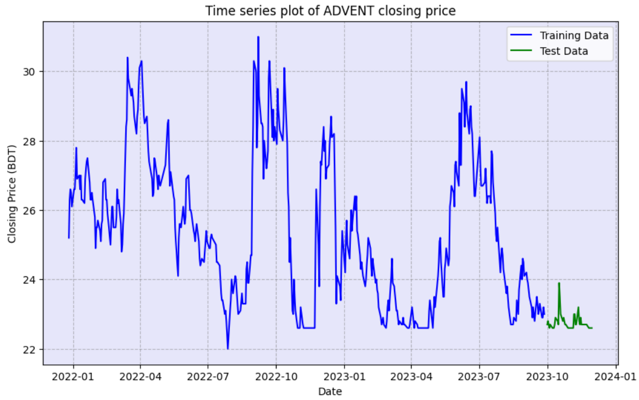
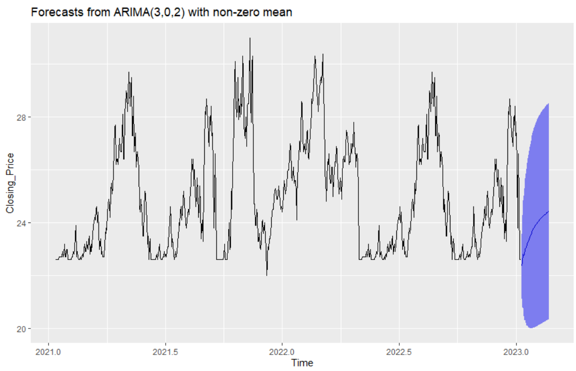
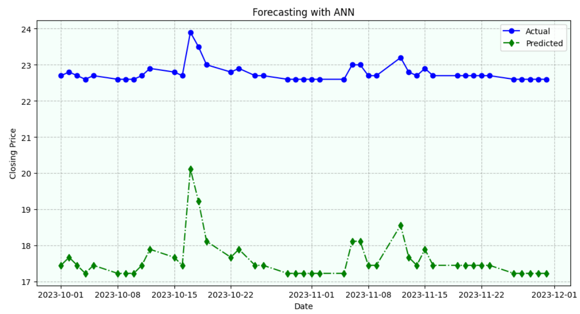


Figure 3.3: Training and Test Dataset of ADVENT

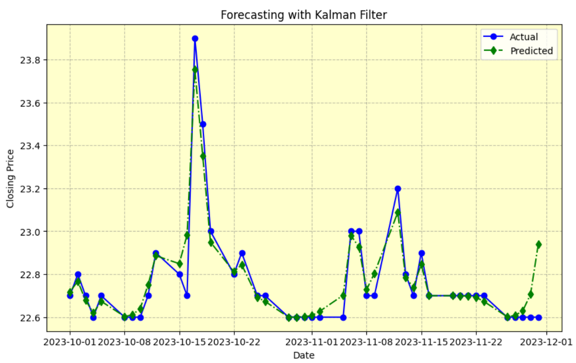
table 3.8, we observe that the Kalman Filter shows less inaccuracy than the other three approaches, making it the optimal forecasting method. Thus, we contrast the forecasting precision of hybrid ARIMA ANN, ANN, and ARIMA. For ADVENT, ARIMA is the most effective forecasting method. Compared to ARIMA, hybrid ARIMA ANN generates slightly more inaccuracy. Out of all the approaches used for ADVENT, ANN shows the largest inaccuracy.



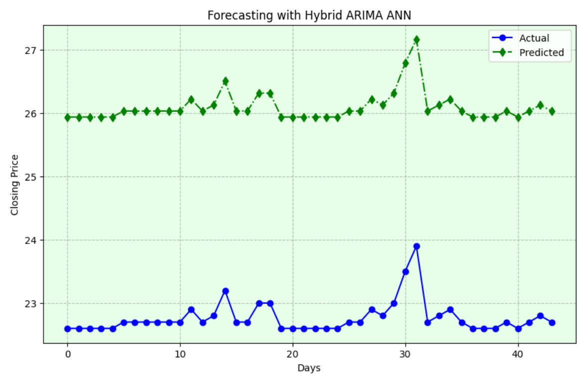
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.4: Actual vs Predicted Closing Price of ADVENT

Table 3.8: Model Performance for ADVENT

Model	RMSE	MAE	MAPE
ARIMA	1.13523397319128	0.995721014637391	0.0438056280233817
ANN	5.174214090548687	5.1653075911782	0.22705533332261005
Kalman Filter	0.0846	0.0482	0.0021
HYBRID ARIMA ANN	3.1584322663418436	3.158344026045366	0.13875756777721315

3.3 ARGONDENIM

Denim fabrics, yarn, cloth, linen, and hosiery goods are manufactured, dyed, finished, and exported by Argon Denims Ltd. The company's headquarters are in Dhaka, Bangladesh, and it was established on July 13, 2006. The company was listed on the DSE in 2013 [32].

Table 3.9: Statistics for Training Dataset of ARGONDENIM

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	430	17.0	21.0	18.431395348837206	18.2	0.5519327405534723

Table 3.10: Statistics for Test Dataset of ARGONDENIM

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	44	18.2	19.3	18.23863636363637	18.2	0.17681406485530354

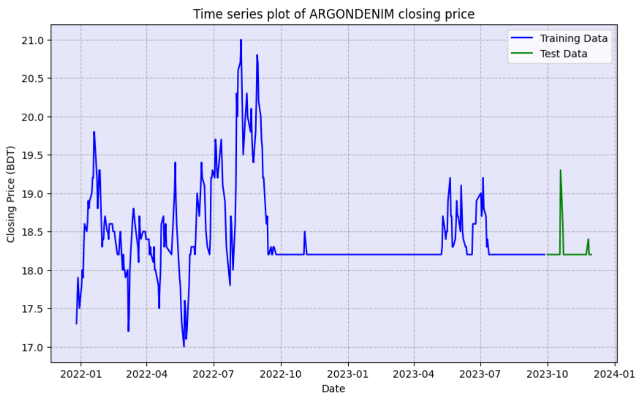


Figure 3.5: Training and Test Dataset of ARGONDENIM

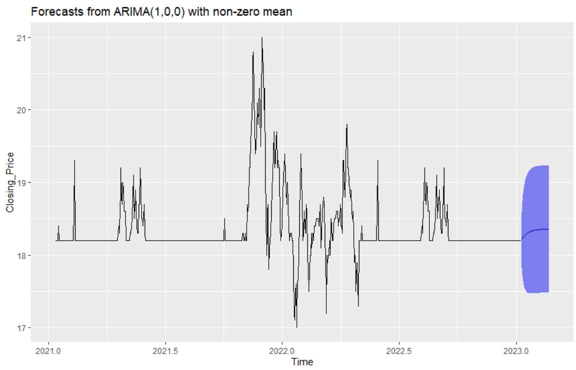
Table 3.11: ADF test, DW test, ARIMA model selection result for ARGONDENIM

Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.01 (stationary)	1.929765(no autocorrelation)	ARIMA(1,0,0)

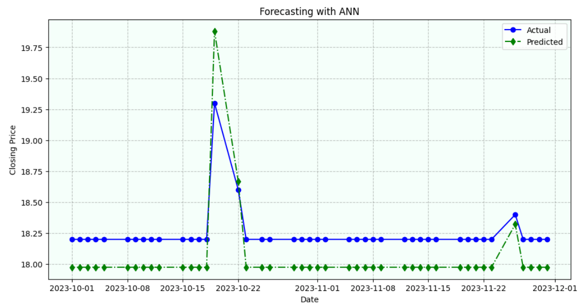
From table 3.11, we observe that the training dataset is stationary. The training datasets exhibit no autocorrelation. Tables 3.9 and 3.10 indicate that the standard deviation between the training and test data is minimal. A modest standard deviation between the training and test data indicates that the model has successfully learnt the patterns in the training data, which accurately represent the overall data distribution. As a result, the model is less likely to be affected by minor changes in the data distribution, making it more resilient. From figure 3.5, we observe that training data exhibits a combination of a long term linear and seasonal trend with nonlinear fluctuations . Figure 3.6 graphically represents the outcomes of ARIMA, ANN, KF, and hybrid ARIMA ANN. From table 3.12, we observe that the Kalman Filter shows less inaccuracy than the other three approaches, making it the optimal forecasting method. Thus, we contrast the forecasting precision of hybrid ARIMA ANN, ANN, and ARIMA. For ARGONDENIM, every forecasting technique is highly effective. They all show extremely low error rates. ARIMA has the minimum error, making it the ideal method for forecasting.

Table 3.12: Model Performance for ARGONDENIM

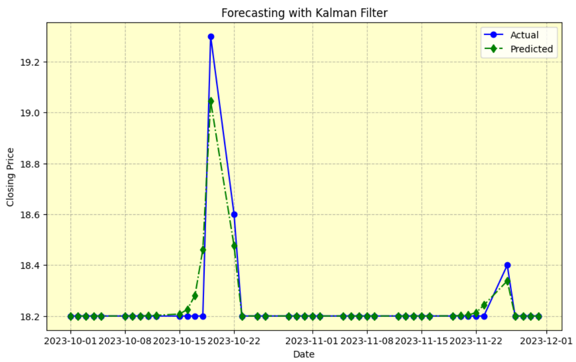
Model	**RMSE**	**MAE**	**MAPE**
ARIMA	0.196558780618329	0.148950321508378	0.00810816318975877
ANN	0.2345854143426307	0.22608378150246286	0.012377996251501764
Kalman Filter	0.0608	0.0201	0.0011
HYBRID ARIMA ANN	0.4129399466164108	0.4128510301763368	0.02264233470512937



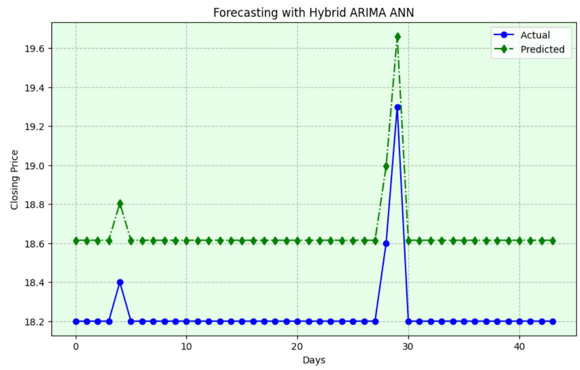
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.6: Actual vs Predicted Closing Price of ARGONDENIM

3.4 BDTHAIFOOD

BD Thai Food & Beverage Ltd is a company that produces a variety of carbonated soft drinks, energy drinks, mango fruit drinks, confections, and spices. The company's headquarters are in Dhaka, Bangladesh, and it was established in 2010 [33].

Table 3.13: Statistics for Training Dataset of BDTHAIFOOD

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	409	11.0	54.1	38.295599022004886	36.9	5.810435186327171

Table 3.14: Statistics for Test Dataset of BDTHAIFOOD

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	44	34.8	34.8	34.8	34.8	0.0

Figure 3.7: Training and Test Dataset of BDTHAIFOOD

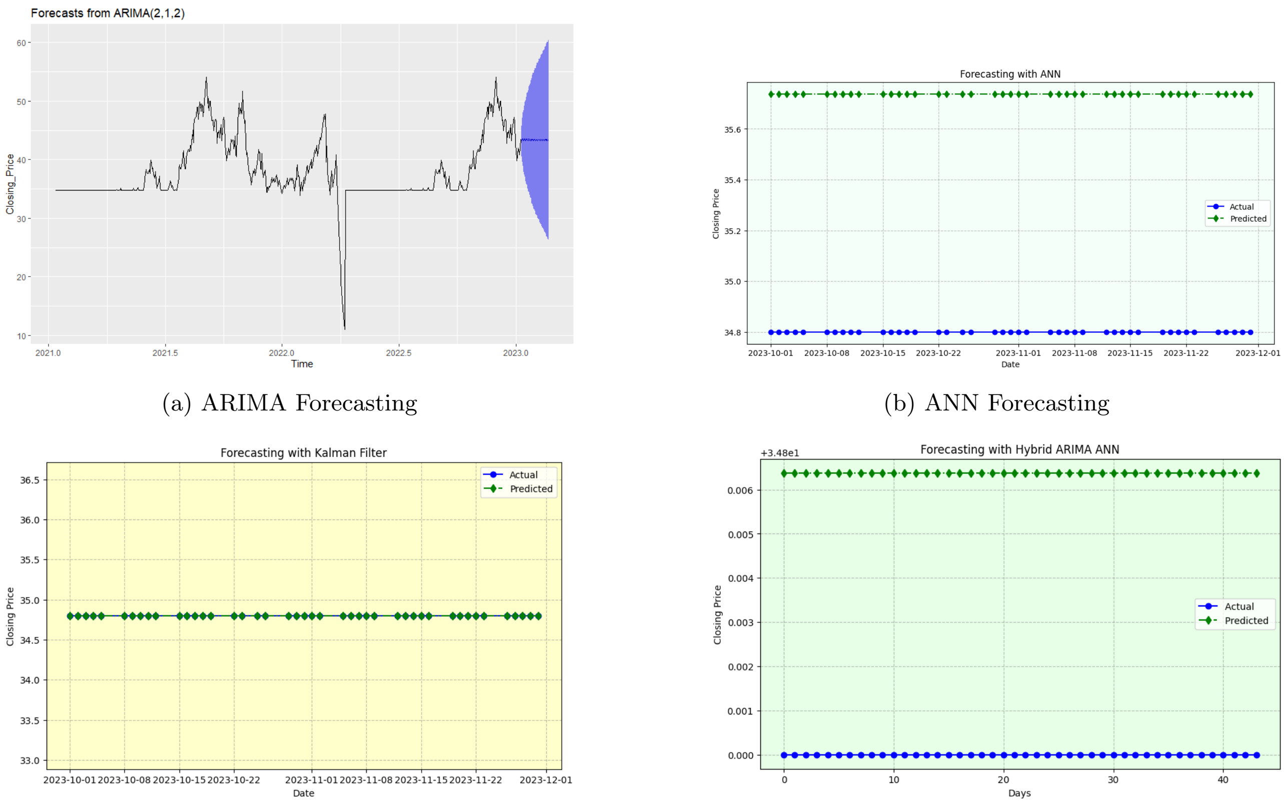
Table 3.15: ADF test, DW test, ARIMA model selection result for BDTHAIFOOD

Augmented Dickey Fuller test (p value)	**Durbin Watson test Statistics**	**ARIMA (p,d,q)**
0.04242 (stationary)	1.996878(no autocorrelation)	ARIMA(2,1,2)

Table 3.16: Model Performance for BDTHAIFOOD

Model	**RMSE**	**MAE**	**MAPE**
ARIMA	8.5774167773248	8.57664697607854	0.24645537287582
ANN	0.9379035949707059	0.937903594970706	0.026951252729043267
Kalman Filter	0.0000	0.0000	0.0000
HYBRID ARIMA ANN	0.006377410888674717	0.006377410888674717	0.00018325893358260684

We can see that the training dataset is stationary from table 3.15. The training datasets do not exhibit autocorrelation. The training dataset's standard deviation is greater than the test dataset's, as can be seen in tables 3.13 and 3.14. A mismatch in the variability or dispersion of the two datasets, caused by random fluctuations, is indicated when the standard deviation of the training data is higher than the standard deviation of the test data. Figure 3.7 shows that the training data combines long-term nonlinear oscillations with a short-term linear trend. Figure 3.8 graphically depicts the outputs of ARIMA, ANN, KF, and hybrid ARIMA ANN. The Kalman Filter is the best forecasting technique since it exhibits less error than the other three methods, as seen in table 3.16. As a result, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN. Hybrid ARIMA ANN is the best for BDTHAIFOOD. A little more error is produced by ANN than by hybrid ARIMA ANN. Due to its enormous error rate, ARIMA is not a good fit for it.



(a) ARIMA Forecasting (b) ANN Forecasting

(c) KF Forecasting (d) Hybrid ARIMA ANN Forecasting

Figure 3.8: Actual vs Predicted Closing Price of BDTHAIFOOD

3.5 ECABLES

Eastern Cables Ltd. is a cable and conductor manufacturing company. The company's headquarters are in Chittagong, Bangladesh, where it was established in 1967.The company was listed on the DSE in 1986 [32].

Table 3.17: Statistics for Training Dataset of ECABLES

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	430	121.8	239.0	174.70116279069768	183.0	27.714679717403182

Table 3.18: Statistics for Test Dataset of ECABLES

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	44	181.3	187.9	181.93636363636367	181.3	1.5362429115534058

Table 3.19: ADF test, DW test, ARIMA model selection result for ECABLES

Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.5033 (stationary)	2.001075(no autocorrelation)	ARIMA(5,1,0)

We can see that the training dataset is stationary from table 3.19. The training datasets do not exhibit autocorrelation. The training dataset's standard deviation is greater than the test dataset's, as can be seen

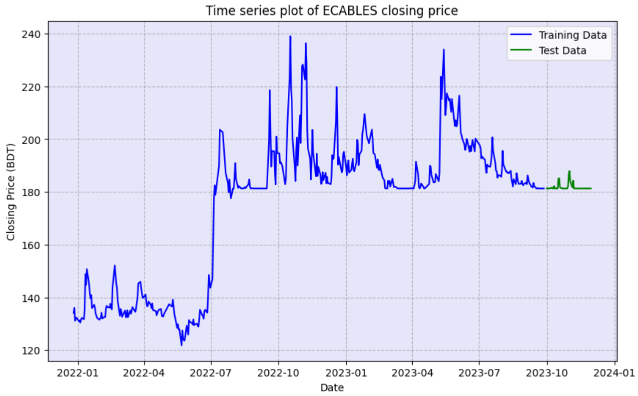


Figure 3.9: Training and Test Dataset of ECABLES

in tables 3.17 and 3.18. A mismatch in the variability or dispersion of the two datasets, caused by random fluctuations, is indicated when the standard deviation of the training data is higher than the standard deviation of the test data. Figure 3.9 shows that the training data demonstrates nonlinear oscillations together with both upward and downward movement. Figure 3.10 shows the results of the hybrid ARIMA ANN, ANN, KF, and ARIMA. The Kalman Filter is the best forecasting technique since it exhibits less error than the other three methods, as seen in table 3.20. As a result, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN. For ECABLES, the optimal forecasting technique is ANN. It produces the least amount of error. ARIMA and Hybrid ARIMA ANN both produce large amounts of error.The hybrid ARIMA-ANN exhibits the largest error. ANN is the only suitable forecasting method for it.

Table 3.20: Model Performance for ECABLES

Model	**RMSE**	**MAE**	**MAPE**
ARIMA	6.79890821896006	6.62423915645368	0.0364804843613028
ANN	0.640747425895723	0.619524314186797	0.003412712641142095
Kalman Filter	0.2850	0.1361	0.0007
HYBRID ARIMA ANN	10.4812971290499	10.481196594238273	0.05761109371627886

3.6 KEYACOSMET

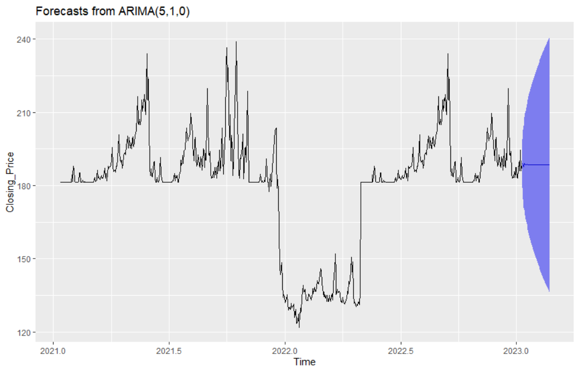
Keya Cosmetics Ltd. produces cosmetics and hygienic products. Established on July 14, 1996, the organisation was domiciled in Banani, Bangladesh, and was founded by Abdul Khaleque Pathan . It was listed on the DSE in 2007 [32].

Table 3.21: Statistics for Training Dataset of KEYACOSMET

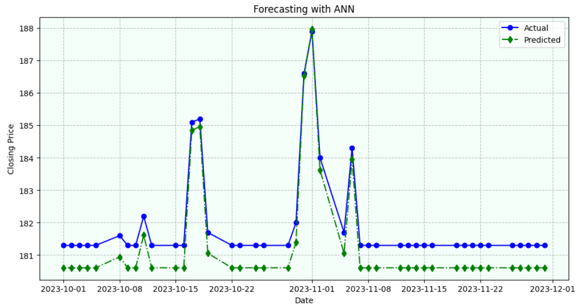
Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	430	6.3	8.4	6.739767441860465	6.5	0.4624097637523316

Table 3.22: Statistics for Test Dataset of KEYACOSMET

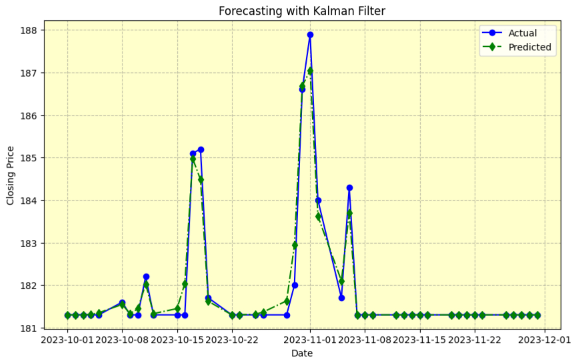
Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	44	6.4	6.6	6.4045454545454525	6.4	0.030151134457776257



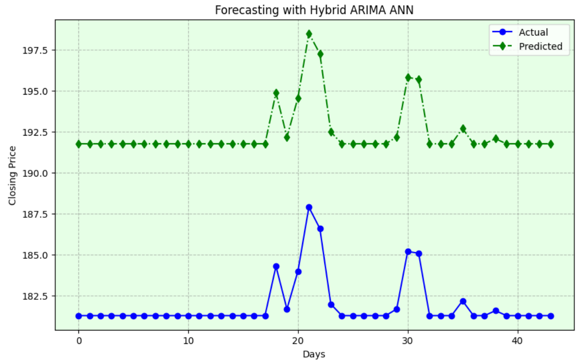
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.10: Actual vs Predicted Closing Price of ECABLES

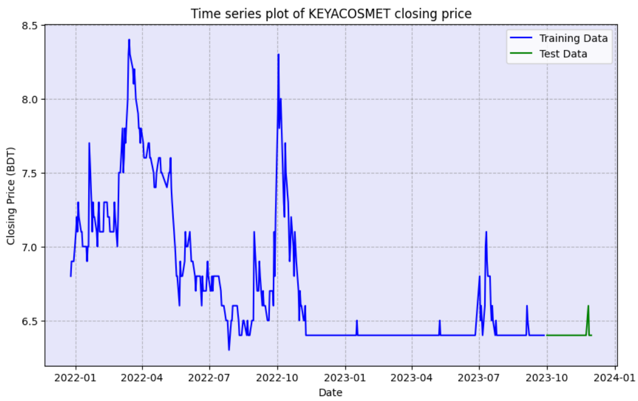
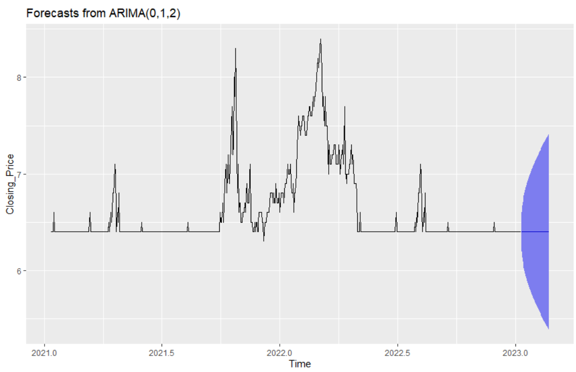


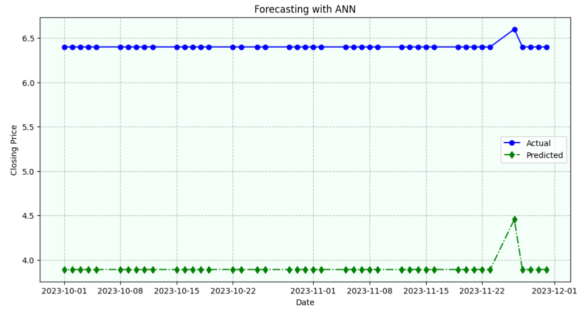
Figure 3.11: Training and Test Dataset of KEYACOSMET

Table 3.23: ADF test, DW test, ARIMA model selection result for KEYACOSMET

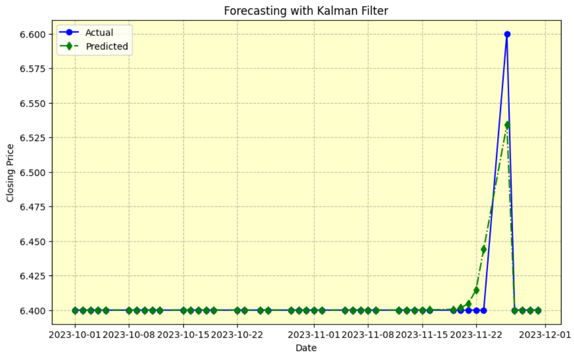
Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.2456 (stationary)	2.006681(no autocorrelation)	ARIMA(0,1,2)



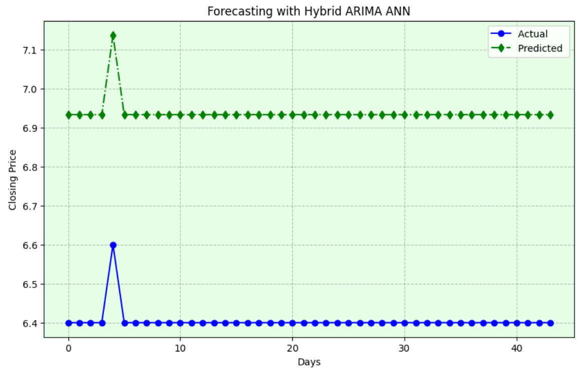
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.12: Actual vs Predicted Closing Price of KEYACOSMET

Table 3.24: Model Performance for KEYACOSMET

Model	**RMSE**	**MAE**	**MAPE**
ARIMA	0.03015113445777763	0.004545454545454545453	0.0006887052341597777
ANN	2.5021536188397198	2.501563821055673	0.3906384874032631
Kalman Filter	0.0122	0.0030	0.0005
HYBRID ARIMA ANN	0.5332658172557728	0.5332655083049425	0.08326494580868514

We can see that the training dataset is stationary from table 3.23. The training datasets do not exhibit autocorrelation. Figure 3.11 shows that the training data demonstrates a combination of nonlinear fluctuation, upward and downward movement, and linear trend. A graphic representation of the results from ARIMA, ANN, KF, and hybrid ARIMA ANN is shown in Figure 3.12. The Kalman Filter is the best forecasting technique since it exhibits less error than the other three methods, as seen in table 3.24. As a result, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN. ARIMA is the most accurate forecasting method for KEYACOSMET. The hybrid ARIMA ANN produces a somewhat higher inaccuracy than ARIMA. In this scenario, the ANN produces a substantial inaccuracy.

3.7 MEGHNAPET

The company owns and runs an industrial plant that processes integral mineral water, manufactures PET bottles, fills edible oil bottles, and sells edible oil and mineral water. On September 17, 1995, the corporation was incorporated as a public company, and on December 21, 1997, it started conducting business. In 2001, the company was listed on the DSE [34].

Table 3.25: Statistics for Training Dataset of MEGHNAPET

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	430	17.9	49.2	33.444186046511625	34.45	7.339807557215723

Table 3.26: Statistics for Test Dataset of MEGHNAPET

Statistics	**Observations**	**Min**	**Max**	**Mean**	**Median**	**Standard Deviation**
Value	44	33.1	42.8	36.37727272727273	35.7	2.610877392965699

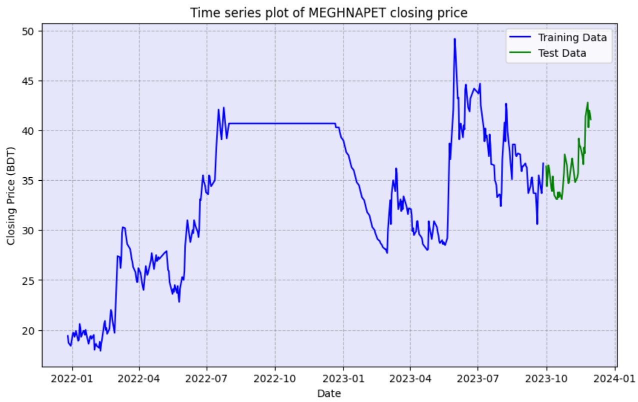


Figure 3.13: Training and Test Dataset of MEGHNAPET

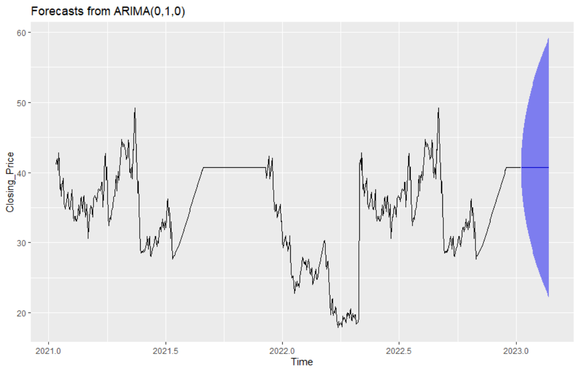
Table 3.27: ADF test, DW test, ARIMA model selection result for MEGHNAPET

Augmented Dickey Fuller test (p value)	**Durbin Watson test Statistics**	**ARIMA (p,d,q)**
0.2266 (stationary)	1.894266(no autocorrelation)	ARIMA(0,1,0)

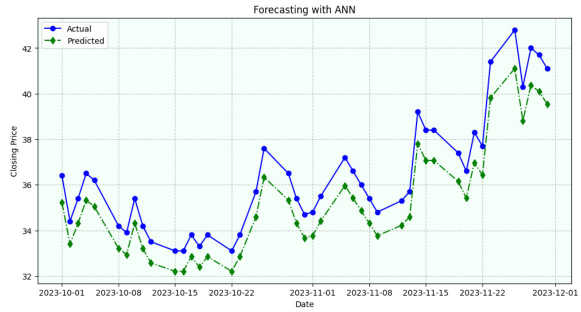
Table 3.28: Model Performance for MEGHNAPET

Model	**RMSE**	**MAE**	**MAPE**
ARIMA	5.03465264668053	4.57272727272727	0.130124422591445
ANN	1.1848623966281557	1.1656645341352982	0.031803173190224716
Kalman Filter	0.6902	0.3363	0.0089
HYBRID ARIMA ANN	1.3658632877216272	1.3603428580544212	0.03770153106965763

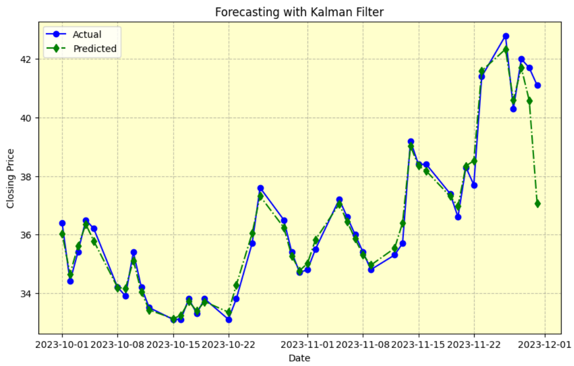
We can see that the training dataset is stationary from table 3.27. The training datasets do not exhibit autocorrelation. It is evident from tables 3.25 and 3.26 that the training dataset's standard deviation is greater than the test dataset's. The presence of random fluctuations in the training data might lead to a mismatch in the variability or dispersion of the two datasets, as shown by a higher standard deviation of the training data than the test data. From figure 3.13, we note that training data exhibits a combination of a short -term linear trend and long -term upward and downward movement with nonlinear variations . Figure 3.14 graphically depicts the outputs of ARIMA, ANN, KF, and hybrid ARIMA ANN. The Kalman Filter is the best forecasting technique since it exhibits less inaccuracy than the other three methods, as seen in table 3.28. As a result, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN.Both ANN and HYBRID ARIMA ANN work incredibly well for MEGHNAPET. The least amount of error is produced by ANN. In this case, ARIMA displays a significant inaccuracy.



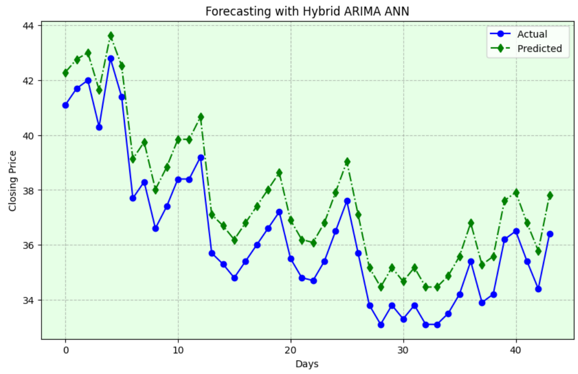
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.14: Actual vs Predicted Closing Price of MEGHNAPET

3.8 PRIMETEX

The company Prime Textile Spinning Mills Ltd. produces and sells mixed and knitted cotton yarn of high quality for export. The company's headquarters are in Dhaka, Bangladesh, and it was established on June 22, 1989 [32].

Table 3.29: Statistics for Training Dataset of PRIMETEX

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	430	21.2	45.0	31.928372093023253	29.2	6.382835713798057

Table 3.30: Statistics for Test Dataset of PRIMETEX

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	44	29.2	29.2	29.200000000000003	29.2	$3.593786877750014e-15$

Table 3.31: ADF test, DW test, ARIMA model selection result for PRIMETEX

Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.9227 (stationary)	2.055521(no autocorrelation)	ARIMA(0,1,0)

We can see that the training dataset is stationary from table 3.31. The training datasets do not exhibit autocorrelation. The training dataset's standard deviation is greater than the test dataset's, as can be seen in

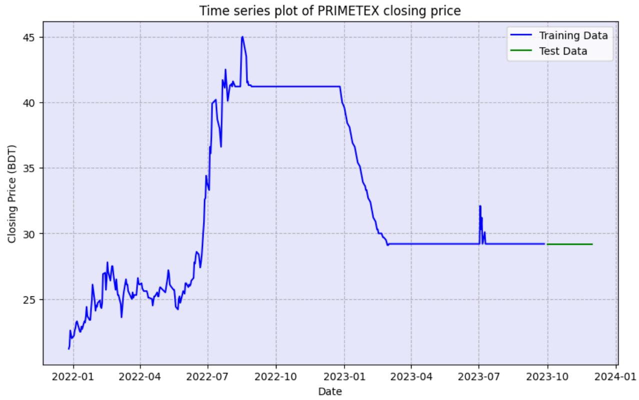
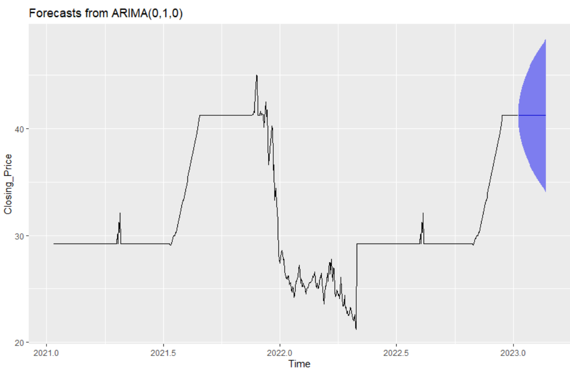
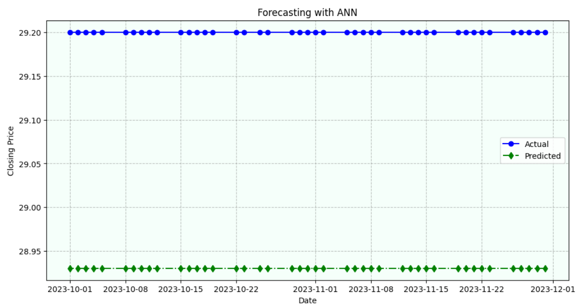


Figure 3.15: Training and Test Dataset of PRIMETEX

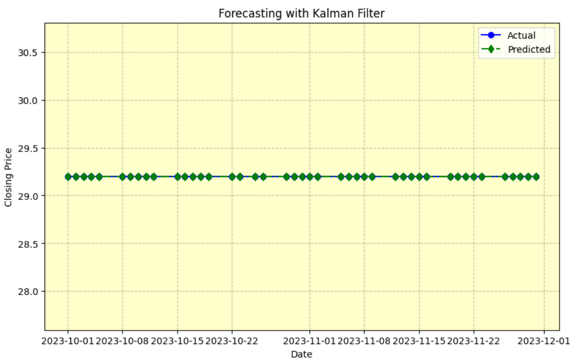
tables 3.29 and 3.30. The presence of random fluctuations in the training data might lead to a mismatch in the variability or dispersion of the two datasets, as shown by a higher standard deviation of the training data than the test data. Figure 3.15 shows that the training data demonstrates both short-term nonlinear fluctuations and a long-term linear trend. An illustration of the results of ARIMA, ANN, KF, and hybrid ARIMA ANN is shown in Figure 3.16. The Kalman Filter is the best forecasting technique since it exhibits less error than the other three methods, as seen in table 3.32. As a result, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN. For PRIMETEX, ANN and hybrid ARIMA ANN function quite effectively. Because of its smaller error, the hybrid ARIMA ANN is the best. ARIMA is inappropriate in this situation because it produces high error rates.



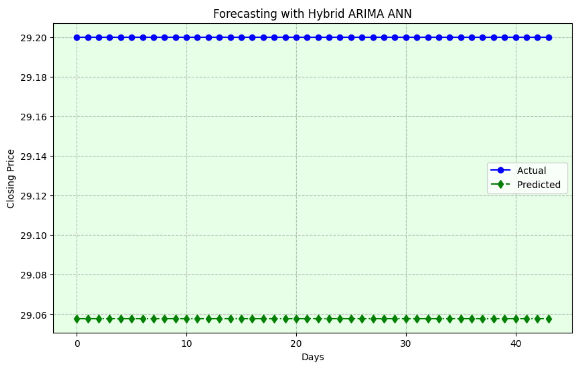
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.16: Actual vs Predicted Closing Price of PRIMETEX

Table 3.32: Model Performance for PRIMETEX

Model	RMSE	MAE	MAPE
ARIMA	12	12	0.410958904109589
ANN	0.26960296630859304	0.26960296630859304	0.009232978298239489
Kalman Filter	0.0000	0.0000	0.0000
HYBRID ARIMA ANN	0.14211387634277273	0.14211387634277273	0.004866913573382627

3.9 UNIONCAP

Financial and investment banking services are offered by Union Capital Ltd. The business was founded on October 9, 1995, and its main office is in Dhaka, Bangladesh. It was listed on the DSE in 2007 [32].

Table 3.33: Statistics for Training Dataset of UNIONCAP

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	430	6.7	13.5	8.744651162790698	8.6	1.2495618790936123

Table 3.34: Statistics for Test Dataset of UNIONCAP

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	44	7.3	8.1	7.506818181818184	7.3	0.26360446697764983

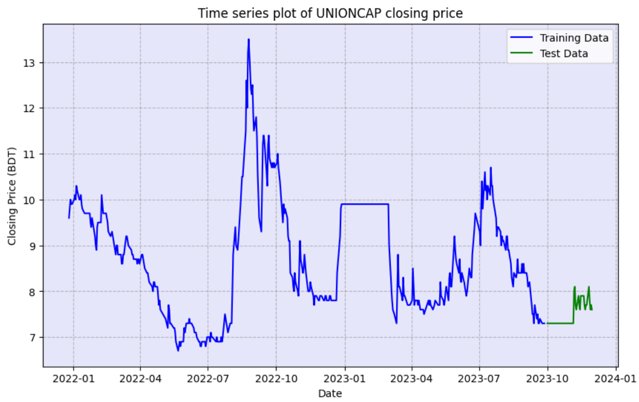
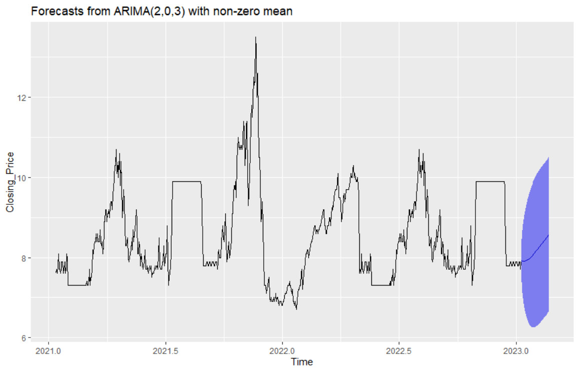


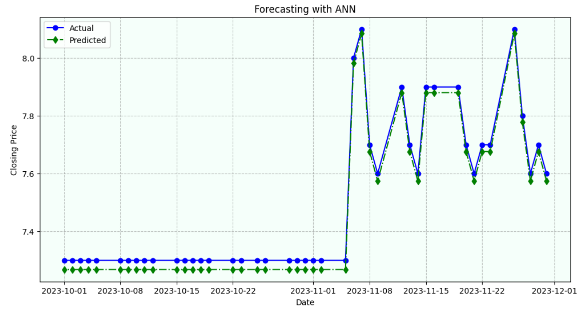
Figure 3.17: Training and Test Dataset of UNIONCAP

Table 3.35: ADF test, DW test, ARIMA model selection result for UNIONCAP

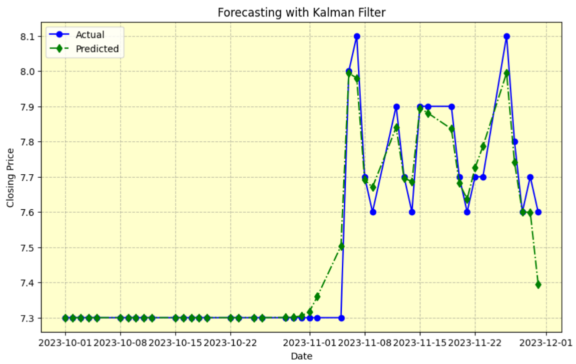
Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.01237 (stationary)	1.998113(no autocorrelation)	ARIMA(2,0,3)



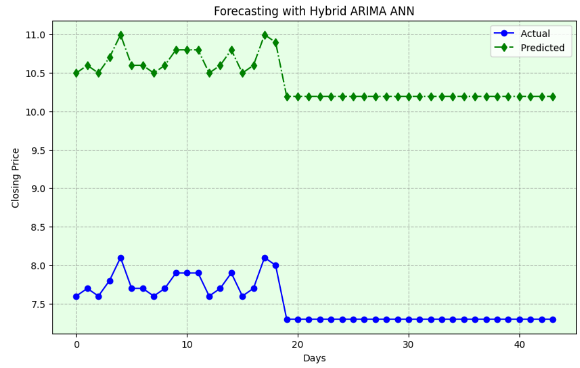
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.18: Actual vs Predicted Closing Price of UNIONCAP

Table 3.36: Model Performance for UNIONCAP

Model	**RMSE**	**MAE**	**MAPE**
ARIMA	0.806929774125817	0.680034734880507	0.0925195273753139
ANN	0.028069802758484218	0.027551278201016407	0.0036986763856192306
Kalman Filter	0.0597	0.0311	0.0041
HYBRID ARIMA ANN	2.8936941438578976	2.8936936031688347	0.3859216407468033

We can see that the training dataset is stationary from table 3.35. The training datasets do not exhibit autocorrelation. We can see that the training data shows intricate or nonlinear patterns in figure 3.17. It combines nonlinear fluctuations with both upward and downward movement. Figure 3.18 graphically depicts the outputs of ARIMA, ANN, KF, and hybrid ARIMA ANN. The Kalman Filter is the best forecasting technique since it exhibits less error than the other three methods, as seen in table 3.36. As a result, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN. The best forecasting technique for UNIONCAP is ANN since it produces the least amount of error. ARIMA exhibits slightly higher inaccuracy than ANN. The most inaccurate model is the hybrid ARIMA ANN.

3.10 PUBALIBANK

Pubali Bank Ltd. provides commercial banking. It offers retail, business, and personal loans. Savings, current, and special notice deposits, demand draft, telegraphic transfer, payment order, teller transaction, locker service, recurring deposits, and term deposits are available. It is headquartered in Dhaka, Bangladesh, and formed in 1959. It was listed on the DSE in 1984 [32].

Table 3.37: Statistics for Training Dataset of PUBALIBANK

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	430	23.9	30.5	26.71418604651163	26.4	1.0292499193788858

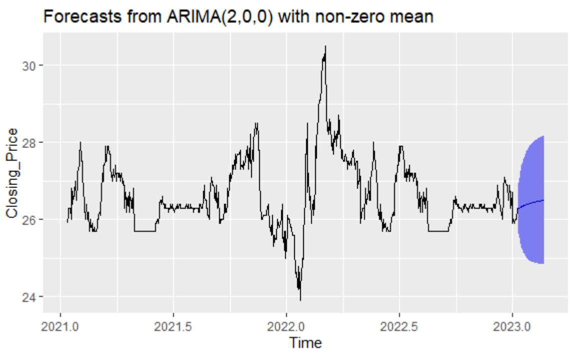
Table 3.38: Statistics for Test Dataset of PUBALIBANK

Statistics	Observations	Min	Max	Mean	Median	Standard Deviation
Value	44	25.7	28.0	26.490909090909096	26.35	0.6141048247077913

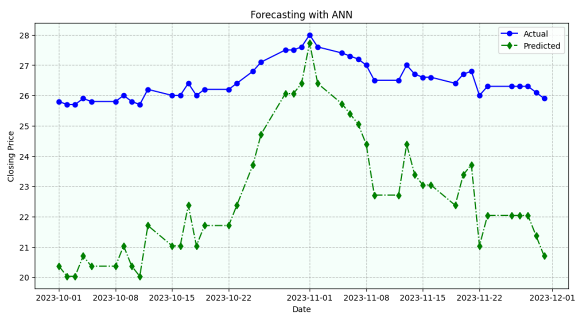
Figure 3.19: Training and Test Dataset of PUBALIBANK

Table 3.39: ADF test, DW test, ARIMA model selection result for PUBALIBANK

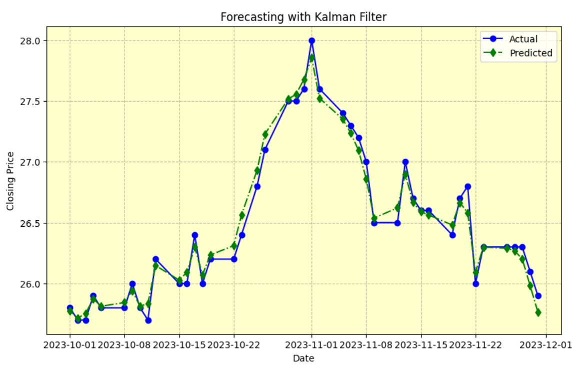
Augmented Dickey Fuller test (p value)	Durbin Watson test Statistics	ARIMA (p,d,q)
0.01 (stationary)	1.99296(no autocorrelation)	ARIMA(2,0,0)



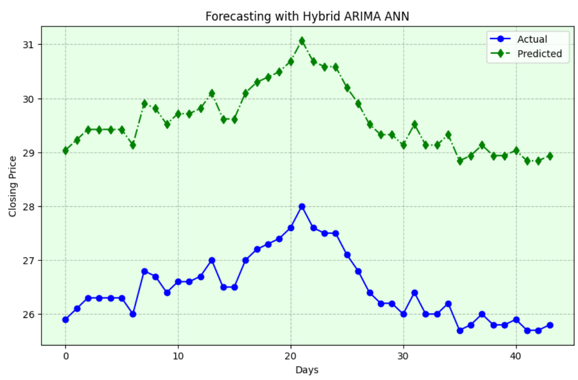
(a) ARIMA Forecasting



(b) ANN Forecasting



(c) KF Forecasting



(d) Hybrid ARIMA ANN Forecasting

Figure 3.20: Actual vs Predicted Closing Price of PUBALIBANK

Table 3.40: Model Performance for PUBALIBANK

Model	RMSE	MAE	MAPE
ARIMA	0.626450943122683	0.506599921144982	0.018981933644331
ANN	4.068898064071023	3.810758842121471	0.14514284771332978
Kalman Filter	0.0877	0.0725	0.0027
HYBRID ARIMA ANN	3.116954548908312	3.116902490095659	0.11773563497116593

We can see that the training dataset is stationary from table 3.39. The training datasets do not exhibit autocorrelation. We can see from figure 3.19 that the data shows an upward and downward movement with nonlinear fluctuations throughout time. An illustration of the results of ARIMA, ANN, KF, and hybrid ARIMA ANN is shown in figure 3.20. The Kalman Filter is the best forecasting technique since it exhibits less error than the other three methods, as seen in table 3.40. As a consequence, we compare the forecast accuracy of ARIMA, ANN, and hybrid ARIMA ANN. For PUBALIBANK, ARIMA is the best forecasting method. ANN and hybrid ARIMA ANN generate significant errors. The ANN generates the largest amount of error.

4 Conclusion

This study employed the ARIMA, ANN, KF, and a hybrid ARIMA ANN model to forecast the closing prices of ten companies listed on the DSE. Evaluation of the model's performance was conducted using RMSE, MAE, and MAPE. It is observed that every approach possesses its own advantages and disadvantages, and the effectiveness of each method might differ depending on the features of the data and the particular trends of the observed stock prices. The key discovery was that when data has a linear trend or seasonal patterns, ARIMA performs better. When the data exhibits a distinct trend or seasonality that the model's parameters (autoregressive, differencing, and moving average terms) can capture, ARIMA performs well. It works efficiently with stationary time series data, or data whose statistical characteristics, such as mean and variance, don't change over time. ARIMA can produce reliable projections when stock prices show seasonal patterns or a distinct linear trend. ARIMA has shown strong performance in capturing distinct trends or seasonality patterns, as observed in the ADVENT, ARGONDENIM, KEYACOSMET, and PUBALIBANK datasets. Data with nonlinear patterns or complicated relationships works well with artificial neural networks. Neural networks are excellent at capturing intricate nonlinear correlations in data, particularly when it comes to deep learning models like ANN. When there are nonlinear relationships or extremely complicated data, they may perform better than other models since they can learn from complex patterns. If there are multiple input features to take into account or if the correlations between the input variables and stock prices are nonlinear, ANN may perform better for stock price forecasting. ANN has shown exceptional proficiency in the cases of ECABLES, MEGHNAPET, and UNIONCAP. The hybrid ARIMA ANN model may be more accurate if there are trends and complex nonlinear interactions.The advantages of both ARIMA and ANN are combined in the hybrid ARIMA ANN. Whereas ANN deals with complicated patterns and nonlinear interactions, ARIMA extracts the data's linear components and temporal patterns. When there are both linear and nonlinear elements in the data, hybrid models can be especially useful. For instance, a hybrid ARIMA ANN model might be more accurate if the stock prices show both straight-line trends and wavy lines when different outside factors are taken into account. The hybrid ARIMA ANN model performed well for AAMRATECH, BDTHAIFOOD, and PRIMETEX. Kalman Filter can estimate system states from noisy data, adapt to dynamic changes, incorporate past information, provide optimal estimates, and conduct prediction and smoothing computationally efficiently, making it valuable for forecasting. Because of its flexibility in changing circumstances, the Kalman Filter is typically able to perform well in stock price forecasting when the data shows a lot of noise or abrupt fluctuations. Among the four forecasting techniques, the Kalman Filter produces the smallest error and performs exceptionally well for the 10 DSE companies. The particulars of the data, such as its trend, seasonality, noise level, and the presence or absence of linear or nonlinear patterns, determine the best forecasting technique. The optimal approach, however, could vary for every company or dataset since different stocks may display distinct behaviours.

References

[1] Jonathan D Cryer & Kung Silk Chan, Time Series Analysis With Applications in R, Second Edition, *Springer*, (2008)

[2] Ruey S Tsay, Analysis of Financial Time Series, Third Edition, *Wiley*, 2010.

[3] Polasek, W. Time Series Analysis and Its Applications: With R Examples, Third Edition by Robert H. Shumway, David S. Stoffer.(2013)

[4] Zhou, Y. . Stock Price Forecasting and Analysis Algorithm Based on ARIMA Taking Shanghai Stock Exchange 50 Index as an Example. *Advances in Economics, Management and Political Sciences*, 48(1), 247–255.(2023)

[5] Rikota, V.,Kovpak, E. . Forecasting the dynamics of the PFTS stock index by using the ARIMA-GARCH model *Pryazovskyi Economic Herald*, 6(17).(2019)

[6] Amjad A. Alsuwaylimi. Comparison of ARIMA, ANN and Hybrid ARIMA-ANN Models for Time Series Forecasting. *Information Sciences Letters*, 12(2), 1003–1016.(2023)

[7] E. H. A. Rady, H. Fawzy and A. M. A. Fattah. (2021). Time series forecasting using tree based methods, *Journal of Statistics Applications and Probability*, 10 (2021) 229–244.(2023)

[8] Govaers, F. (Ed.). . Introduction and Implementations of the Kalman Filter. *IntechOpen.*(2019)

[9] Wei, L.-Y. A hybrid model based on ANFIS and adaptive expectation genetic algorithm to forecast TAIEX. Economic Modelling, 33:893–899.(2013).

[10] D. C. Montgomery, C. L. Jennings and M. Kulahci, Introduction to time series analysis and forecasting *John Wiley & Sons.*(2015)

[11] P. B. Patel and T. Marwala, ”Forecasting closing price indices using neural networks,” *2006 IEEE International Conference on Systems, Man and Cybernetics*, Taipei, Taiwan,, pp. 2351-2356,(2006)

[12] Zhang, G., Patuwo, B., & Hu, M. Y. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research*, 28(4), 381–396.(2001)

[13] Khashei, M., & Bijari, M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 11(2), 2664–2675.(2011)

[14] Büyükşahin, M. A., & Ertekin, E. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing*, 361, 151–163. (2019)

[15] R. H. Shumway, D. S. Stoffer and D. S. Stoffer, Time series analysis and its applications *Springer (2000)*

[16] H. Fawzy, E. H. A. Rady and A. M. A. Fattah, Forecasting time series using a hybrid ARIMA-ANN methodology, *Journal of Applied Probability and Statistics* 16 (2021) 95–106.

[17] M. Khashei and M. Bijari, An artificial neural network (p, d, q) model for timeseries forecasting, *Expert Systems with Applications* 37(1) 479–489.(2010)

[18] L. Wang, H. Zou, J. Su, L. Li and S. Chaudhry, An ARIMA-ANN Hybrid Model for Time Series Forecasting, *Systems Research and Behavioral Science* 30(3) 244–259.(2013)

[19] Sen, J. &Datta Chaudhuri, T. Understanding the sectors of Indian economy for portfolio choice. *International Journal of Business Forecasting and Marketing Intelligence*, 4(2), 178-222.(2018a)

[20] Sen, J. & Datta Chaudhuri, T. Stock price prediction using machine learning and deep learning frameworks.*Proceedings of the 6th International Conference on Business Analytics and Intelligence* (ICBAI'18), Bangalore, India,December 20-22, (2018b)

[21] Bepari, M.K. and Mollik, A. Bangladesh Stock Market Growing? a key indicators based assessment. *Journal of Business Administration Online (JBAO)*, Issue 8, 2008, Arkansas: School of Business, Arkansas Tech Universi. (2008)

[22] Plummer, T. Forecasting financial markets. *Financial Times.*(2006).

[23] P. G. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing* 50 (2003) 159–175.

[24] Sen, J. & Datta Chaudhuri, T. A time series analysis-based forecasting framework for the Indian healthcare sector.*Journal of Insurance and Financial Management*, 3(1), 66-94. (2017a)

[25] Sen, J. & Datta Chaudhuri, T. A predictive analysis of the Indian FMCG sector using time series decomposition based approach. *Journal of Economics Library*, 4(2), 206-226.(2017b)

[26] Sen, J. & Datta Chaudhuri, T. A time series analysis-based forecasting approach for the Indian realty sector.*International Journal of Applied Economic Studies*, 5(4), 8 – 27.(2017c)

[27] Sen, J. & Datta Chaudhuri, T. A robust predictive model for stock price forecasting. *Proceedings of the 5th International Conference on Business Analytics and Intelligence*, Bangalore,India, December 11-13, 2017

[28] Ghahnavieh, A. E. . Time series forecasting of styrene price using a hybrid ARIMA and neural network model. *Independent Journal of Management & Production*, 10(3), 915.(2019)

[29] P. Barba, B. Rosado, J. Ramırez-Zelaya and M. Berrocoso, Comparative Analysis of Statistical and Analytical Techniques for the Study of GNSS Geodetic Time Series, *Engineering Proceedings* 5(1) p. 21 (2021)

[30] H. Fawzy, E. H. A. Rady and A. M. A. Fattah, Comparison between support vector machines and k-nearest neighbor for time series forecasting,*Journal of Mathematical and Computational Science* 10(6) (2020) 2342–2359

[31] Dhaka Stock Exchange Ltd. URL: *https://dsebd.org/*

[32] Market Watch. URL: *https://marketwatch.com/*

[33] Morningstar. URL: *https://morningstar.com/*

[34] EMIS. URL: *https://emis.com/*

[35] Chowdhury, T. U., & Islam, M. S. (2021). ARIMA Time Series Analysis in Forecasting Daily Stock Price of Chittagong Stock Exchange (CSE). *International Journal of Research and Innovation in Social Science*, 05(06), 214–233.

[36] Rana, D., Mir, N. B., Dwivedi, A., Nagpal, R., & Mehrotra, D. (2023). Predicting the future trends for circular economy.

[37] Brownlee, J. (2020). A Gentle Introduction to the Rectified Linear Unit (ReLU). *MachineLearningMastery.com.* https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/

[38] Li, Q., Li, R., Ji, K., & Dai, W. (2015, November). Kalman filter and its application. *In 2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)* (pp. 74-77). IEEE.

[39] Hun, L. C., Yeng, O. L., Sze, L. T., & Chet, K. V. (2016). Kalman filtering and its real-time applications. *Real-time Systems*, 93-116.

[40] Chi, Y. N. (2021). Time Series Forecasting of Global Price of Soybeans using a Hybrid SARIMA and NARNN Model.*Data Science: Journal of Computing and Applied Informatics*, 5(2), 85–101.

[41] Aladag, C. H., Egrioglu, E., & Kadilar, C. (2009). Forecasting nonlinear time series with a hybrid methodology. *Applied Mathematics Letters*, 22(9), 1467–1470.

[42] Hossan, M. S., Islam, M. S., & Kamrujjaman, M. (2022). Efficient Numerical Schemes for Computations of European Options with Transaction Costs. European Journal of Mathematical Analysis, 2, 9.

[43] Al Mobin, M., & Kamrujjaman, M. (2023). Downscaling epidemiological time series data for improving forecasting accuracy: An algorithmic approach. *PLOS ONE*, 18(12), e0295803.

[44] Mohammad, K. M., Tisha, M. S., & Kamrujjaman, M. (2024). Wiener and Lévy processes to prevent disease outbreaks: Predictable vs stochastic analysis. *Partial Differential Equations in Applied Mathematics*, 10, 100712.

APPENDIX A: ADF Test and DW Test Statistic

A.1. ADF Test

The augmented Dickey-Fuller test is used to determine if a time series is stationary or not. It is significant for ARIMA forecasting because the ADF test helps validate the stationarity assumption required for ARIMA models, which perform best with stationary data. When statistical characteristics such as mean, variance, and autocorrelation do not change over time, the time series is said to be stationary. A non-stationary series, on the other hand, exhibits trends, seasonality, or alterations in statistical characteristics with time.

The 'I' (integrated) component of ARIMA works with differencing with the goal of stationaryizing a non-stationary time series. The ADF test finds the existence of a unit root in the data, which aids in determining whether differencing is required.

The p-value in the ADF test indicates the strength of the evidence supporting the series' stationarity. It indicates the probability that the null hypothesis (the series has a unit root and is non-stationary) is true.

A small p-value (less than 5%) suggests strong evidence against the null hypothesis, implying that the data is stationary. This would lead to rejecting the null hypothesis in favour of stationarity.

Conversely, a larger p-value suggests that there isn't enough evidence to reject the null hypothesis, indicating that the data might be non-stationary.

A.2. DW Test Statistic

In statistical analysis, the Durbin-Watson test statistic is frequently used to find autocorrelation in the residuals of a regression or time series model. The Durbin-Watson test can be used in the forecasting context to determine if residual autocorrelation from a forecasting model, like the ARIMA model, still exists.

The presence of autocorrelation in residuals suggests that the model's errors (residuals) at various time lags exhibit certain patterns or correlations. Autocorrelation in residuals in time series forecasting could mean that the model hasn't fully taken into account all the temporal connections that can be seen in the data. This could lead to biassed or inefficient forecasts.

The residuals should be independent and identically distributed (i.i.d.), which is one of the fundamental presumptions of many forecasting models, including ARIMA. The independence assumption has failed due to autocorrelation in residuals. By analysing the residuals' autocorrelation, the Durbin-Watson test assists in determining if this violation has occurred.

The model could not capture some underlying patterns in the data if residuals show considerable autocorrelation. This may have an impact on the forecasts' dependability and accuracy.

The range of the Durbin-Watson test statistic is 0–4. A test statistic score of about 2 indicates that there is no significant autocorrelation. Values near 0 indicate positive autocorrelation, while values near 4 indicate negative autocorrelation. The formula for the Durbin-Watson test statistic involves calculating the sum of squared differences between adjacent residuals in a regression analysis and then normalising it.

When the Durbin-Watson test statistic shows that a forecasting model's residuals are autocorrelation-free, it usually means that the model accurately represents the temporal dependencies found in the data. The absence of autocorrelation in the residuals, which confirms the assumptions of independence and randomness, supports the forecasting model's ability to accurately capture the temporal patterns found in the data.

APPENDIX B: Implementation of the Kalman Filter

Let

$$\begin{aligned}
\hat{\underset{\sim}{x}}_{t|t-1} &= \mathbb{E}[x_t|\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_{t-1}], \\
P_{t|t-1} &= \text{var}\{\underset{\sim}{x}_t - \hat{\underset{\sim}{x}}_{t|t-1}\}, \\
\hat{\underset{\sim}{x}}_{t|t} &= \mathbb{E}[x_t|\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_t], \\
P_{t|t} &= \text{var}\{\underset{\sim}{x}_t - \hat{\underset{\sim}{x}}_{t|t}\},
\end{aligned}$$

Then we have the following updating equations for the Kalman filter.

$$\begin{aligned}
\hat{\underset{\sim}{x}}_{t|t-1} &= F\hat{\underset{\sim}{x}}_{t-1|t-1} \\
P_{t|t-1} &= FP_{t-1|t-1}F' + Q \\
\hat{\underset{\sim}{x}}_{t|t} &= \hat{\underset{\sim}{x}}_{t|t-1} + K_t(\underset{\sim}{y}_t - G\hat{\underset{\sim}{x}}_{t|t-1}) \\
P_{t|t} &= P_{t|t-1} - K_tGP_{t|t-1}
\end{aligned}$$

where

$$K_t = P_{t|t-1}G'(GP_{t|t-1}G' + R)^{-1}$$

is the **'gain'** at time t and

$$\underset{\sim}{y}_t - G\hat{\underset{\sim}{x}}_{t|t-1} = \underset{\sim}{y}_t - \mathbb{E}[\underset{\sim}{y}_t|\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_{t-1}] = \underset{\sim}{y}_t - \hat{\underset{\sim}{y}}_{t|t-1}$$

is the **'innovation'**.

- Iteration 0 (initialisation)

$$\begin{aligned}
\hat{\underset{\sim}{x}}_{0|0} &= \mathbb{E}[\underset{\sim}{x}_0] = 0, \\
P_{0|0} &= \text{var}\{\underset{\sim}{x}_0 - \underset{\sim}{0}\}
\end{aligned}$$

- Iteration 1

$$\begin{aligned}
&\hat{\underset{\sim}{x}}_{1|0} = F\hat{\underset{\sim}{x}}_{0|0} = 0 \\
&P_{1|0} = \text{var}\{\underset{\sim}{x}_1 - \underset{\sim}{0}\} = P_{0|0}(\text{stationarity}) \\
&\hat{\underset{\sim}{x}}_{1|1} = K_1\underset{\sim}{y}_1 = P_{1|0}G'(GP_{1|0}G' + R)^{-1}\underset{\sim}{y}_1 \\
&P_{1|1} = (I - K_1G)P_{1|0}
\end{aligned}$$

- Iteration 2

$$\begin{aligned}
&\hat{\underset{\sim}{x}}_{2|1} = F\hat{\underset{\sim}{x}}_{1|1} \\
&P_{2|1} = FP_{1|1}F' + Q \\
&\hat{\underset{\sim}{x}}_{2|2} = \hat{\underset{\sim}{x}}_{2|1} + K_2(\underset{\sim}{y}_2 - G\hat{x}_{2|1}), \quad \text{where} \quad K_2 = P_{2|1}G'(GP_{2|1}G' + R)^{-1} \\
&P_{2|2} = (I - K_2G)P_{2|1}
\end{aligned}$$

- Iteration $t = 3, 4, \cdots, N$

$$\hat{\underset{\sim}{x}}_{t|t-1} = F\hat{\underset{\sim}{x}}_{t-1|t-1}$$

$$
\begin{aligned}
P_{t|t-1} &= FP_{t-1|t-1}F' + Q \\
\hat{\underset{\sim}{x}}_{t|t} &= \hat{\underset{\sim}{x}}_{t|t-1} + K_t(\underset{\sim}{y}_t - G\hat{x}_{t|t-1}), \quad \text{where} \quad K_t = P_{t|t-1}G'(GP_{t|t-1}G' + R)^{-1} \\
P_{t|t} &= (I - K_tG)P_{t|t-1}
\end{aligned}
$$