A COMPARATIVE ANALYSIS OF ARIMAX AND NNAR IN MODELING AND FORECASTING INFLATION INCORPORATING CLIMATE CHANGE.

A research project submitted to the Department of Statistics and Actuarial Sciences in the School of Mathematics and Physical Sciences in partial fulfillment of the requirement for the award of the degree of BSc. Financial Engineering at Jomo Kenyatta University of Agriculture and Technology

2024

Declaration

This research project is our original work and has not been presented for a degree in any other university.

Name	Registration No.	Signature
Ruth Kariuki Wanjiru	SCM222- 1261/2021	....
Mickerina Wangithi Wainaina	SCM222- 0539/2021	....
Taylor Nanjala Wanyama	SCM222- 1316/2021	....
Audrey Chepkemboi Bii	SCM222-0559/2022	....
Grace Mogaya Meremo	SCM222-1307/2021	....

This research project has been submitted for examination with my approval as the university supervisor.

Name: Dr. Bonface Miya Malenje

Signature: ....

Date: ....

Department of Statistics and Actuarial Sciences,

Jomo Kenyatta University of Agriculture and Technology.

Acknowledgement

We express our heartfelt gratitude to the Almighty for granting us the strength, courage, grace, and empowerment to reach this point in our academic journey and for presenting us with the opportunity to successfully advance this proposal. Our sincere appreciation goes to Dr. Bonface Malenje, our supervisor, for his invaluable assistance, guidance, and patience throughout the research process. We extend our thanks to our parents, who have been a constant source of encouragement during challenging times, and to our friends for their unwavering support and friendship. We would also like to thank the JKUAT fraternity and the department, for equipping us with sufficient knowledge to undertake this project.

Contents

List of Figures

List of Abbreviations

ARIMAX - Autoregressive Integrated Moving Average with eXogenous Variable

CPI - Consumer Pricing Index

ADF - Augmented Dickey-Fuller

MLE - Maximum Likelihood Estimation

ACF - AutoCorrelation Function

PACF - Partial - Autocorrelation Function

MSE - Mean Squared Error

RMSE - Root Mean Squared Error (RMSE)

ML - Machine Learning

KMD - Kenya Meteorological Department

NNAR - Neural Network Autoregressive

ARIMAX-LSTM - Autoregressive Moving Average Exogeneous LongShort-Term Memory

MAE - Mean Absolute Error

ReLU - Rectifier Learning Unit

Chapter 1

Introduction

Climate change has been one of the most critical global challenges in recent years, influencing not only the state of the environment, but also economic stability. Indeed, the relationship between climate change and inflation is complex, as disruptions in agricultural output, volatility in energy prices, and infrastructure damage all drive cost-push inflation. Despite these recognized connections, there is still a lack of comprehensive studies using advanced econometric and machine learning models to analyze the impact of climate change on inflation. This study tries to fill this gap by comparing the performance of the ARIMAX and NNAR models. While ARIMAX incorporates climate-related variables such as anomalies in temperature and fluctuations in energy prices into time series modeling, NNAR is a machine learning approach that can lighten up nonlinear relationships and complex patterns. We analyze these models to provide information to assist policymakers in designing ways through which inflationary pressures could be cushioned in a climate-sensitive world.

1.1 Background of the Study

The linkage of climate change and various economic variables, especially inflation, has drawn extensive interest in recent years, as environmental factors have gained momentum in their influence on economic stability. There are various ways in which climate change feeds into inflationary pressure, such as disruptions in agricultural production, volatility in energy prices, and damage to infrastructure Boneva and Ferrucci (2022). While climate variables, especially temperature anomalies, have been seen as one of the driving forces of inflation, most traditional models of inflation cannot correctly include them Durevall and Ndungu (2001). In light of this, this research tries to bridge this gap by using two advanced econometric models, ARIMAX and NNAR, in forecasting inflation with the inclusion of climate change variables. While the ARIMA model does not include exogenous variables, ARIMAX extends this and makes it wider in scope, to capture those external factors that could even include climate anomalies Bolanle and OLUWADARE (2017). ARIMAX, an extension of the ARIMA model, allows for the inclusion of exogenous variables, thus broadening its scope in capturing external factors such as climate anomalies. But while ARIMAX captures linear relationships, it falls short in dealing with the higher degree of complexity and nonlinearity between the interactions of climate change and inflation. At this juncture, the NNAR models, incorporating machine learning techniques, hold certain advantages. The NNAR class tries to model in such a way as to capture nonlinearities in data, hence being able to provide deeper insights into the intricate, dynamic nature of inflation drivers in the context of climate change Zhang et al. (1998). Previous studies have explored the integration of climate variables into economic models, yet few have undertaken a comparative analysis of ARIMAX and NNAR models in the context of inflation forecasting. For instance, John et al. (2024) developed an integrated ARIMAX-LSTM hybrid model to provide a prediction on inflation, in which the outcomes demonstrate the hybrid model was performing better than a single linear or nonlinear model. This hybridization shows a promising direction for addressing the limitations inherent in traditional models. The failure of the existing models of inflation forecasting, especially in terms of their inability to incorporate the effects of climate change appropriately, calls for innovative approaches that integrate economic and environmental variables. This paper, therefore, seeks to fill this gap by comparing the performance of ARIMAX and NNAR models in forecasting inflation, with a particular focus on the influence of climate-related factors such as temperature anomalies. In this regard, the research could help enhance policymakers', financial institutions', and investors' ability to predict and control one of the main economic risks related to climate-induced inflation.

1.2 Statement of the Problem

Despite the clear economic risks posed by climate change, current inflation models do not fully account for these environmental factors. Traditional models like ARIMA focus on the historical behavior of inflation but cannot incorporate external variables that significantly influence inflation in today's world. Compared to that, ARIMAX improves with the inclusion of exogenous variables but captures mainly linear relationships. However, some of the interactions between climate change and inflation are nonlinear and complex, and traditional models cannot address these. The NNAR model represents a promising alternative with roots in machine learning, directly modeling nonlinear patterns and therefore uncovering intricate relationships in the data.

This project tries to solve the problem that all the traditional inflation forecasting models fail to account for the increasing impact of climate change on inflation dynamics. With increased climate change, considering environmental factors like the current paper, along with traditional economic variables holds the prospect of much better forecasts of inflation. This project compares ARIMAX and NNAR modeling and prediction methodologies that integrate key climatic variables, such as temperature anomalies, toward the forecasting of inflation. In return, the research tries to enhance the capability of the actual policymakers, financial institutions, and investors in regard to foreseeable economic risks from climate-induced inflation. Comparing such models would hint at their respective strengths and weaknesses, offering a complete approach to understanding and mitigating climate-driven inflationary pressures.

1.3 General Objective

To compare ARIMAX and NNAR in modelling and forecasting infation incorporating climate change.

1.3.1 Specific Objectives

1. To fit ARIMAX and NNAR models using CPI incorporating climate change.

2. To forecast inflation using the ARIMAX and NNAR models.

3. To evaluate the performance of the forecasting models.

1.4 Justification of the Study

Climate change is also a major factor affecting most macroeconomic indicators, including inflation. While the traditional models, such as ARIMA, do not have the ability to include external variables, the enhanced capabilities of ARIMAX and NNAR models include these variables. ARIMAX is able to integrate exogenous variables for better accuracy in the forecasting of climate-sensitive variables. On the other hand, NNAR can model nonlinear relationships between climate change and inflation. Comparing these models will provide valuable lessons that can help policymakers anticipate and address climate-induced inflation effectively.

By integrating climate related factors such as temperature changes as the exogeneous Variable, the research aims to quantify its effect on inflation changes in the Kenyan economy. The findings will provide valuable insights for policymakers and financial institutions to come up with strategies which are more sustainable. Additionally, policymakers can forecast future inflation more accurately and design more effective monetary and fiscal policies, hence helping central bank and government institutions better respond to inflation shocks arising from climate change. It also offers more understanding of how environmental factors especially temperature changes drive inflation.

Chapter 2

Literature Review

2.1 Introduction

This section reviews the applicability of the ARIMAX and NNAR models in inflation forecasting, with a specific focus on incorporating climate variables, such as temperature, as exogenous factors. The review encompasses the theoretical framework, empirical studies, and practical applications of these models to provide a foundation for their comparative analysis.

2.2 Theoretical framework

Previous research into Kenya's inflation modeled factors such as money supply and exchange rates alongside food and non-food world prices Durevall and Sjo (2012), but does not consider temperature variations like climate change shocks. Scientists now recognize climate change as a major factor in macroeconomic stability but find this factor missing from modern inflation tools Boneva and Ferrucci (2022). Central banks depend on overall data but they miss important climate information from local areas. Boneva and Ferrucci advocate for a suite-of-models approaches but they fail to propose one unique model. Our research uses ARIMAX incorporating exogenous variables like temperature and NNAR with lagged values from time series data to forecast the impact of climate change on Kenyan inflation

ARIMAX models extend the traditional ARIMA framework by including exogenous variables, enabling a broader analysis of external influences on a dependent variable. According to Bolanle and OLUWADARE (2017), ARIMAX combines univariate time series and regression components, allowing dynamic relationships among variables to be captured. This makes it suitable for incorporating temperature, a key factor in understanding climate-driven inflation changes. Similarly, Neural Network Autoregressive (NNAR) models offer flexibility in modeling nonlinear and complex interactions in time series data. Zhang et al. (1998) noted, NNAR models excel in capturing patterns traditional models might miss, particularly when nonlinear relationships are present. Whether machine learning models truly outperform other methods depends on multiple elements including target indicator specifications, data generation processes, length of forecasts and quality of input data Makridakis et al. (2018). While ARIMAX is rooted in econometrics, NNAR leverages machine learning to address time series forecasting challenges.

2.3 Empirical Studies on Climate Variables in Forecasting

Comparing ARIMAX and NNAR models, large differences can be seen in the strengths and limits set by these two. The ARIMAX models, by statistical theory, hold a place in the scenario when exogenous variables, such as temperature, are completely specified and the effect caused is assumed linear. NNAR models, on the other hand, capture nonlinear and complex interactions, thereby offering flexibility in cases where the relationships between variables are ill-defined or highly dynamic. Hybrid models using ARIMAX combined with neural networks have also been developed. The application of an ARIMAX model to capture even the exogenous variables, like temperature, demonstrates its efficiency through empirical research. Elshewey et al. (2022) used a seasonal ARIMAX model along with wavelet decomposition in temperature forecasting. Though their work was related to temperature prediction, its implication for integrating climate data into inflation forecasting is huge. However, findings by Gathingi (2014) in Kenya showed that traditional ARIMA models were incapable of handling external shocks such as climate variability. Her findings necessitated an ARIMAX model with the inclusion of exogenous variables-temperature, as a factor in the model, complementing the assertions of Durevall and Ndungu (2001) in insisting on modeling dynamic inflation for climate-sensitive economies.

NNAR models have been found capable of estimating performance for complex time series, which are influenced by exogenous variables. Farid et al. (2016) indicated the performance of the model in analyzing temperature-driven agricultural yield predictions, citing appropriateness in nonlinear relationships. Though this study has not channeled its focus on inflation, the principles of NNAR modeling offer a backbone on which to base its application in this study within an economic context. Boneva and Ferrucci (2022) even recommended adding climate change variables into economic models, citing the inability of traditional frameworks to capture the economic flux occasioned by climate change. Such a recommendation agrees with the capability of NNAR to adapt to different and nonlinear data, making them very suitable for inflation forecasting in climate-sensitive scenarios. tried. For example, John et al. (2024)proposed a hybrid model of ARIMAX-LSTM for inflation forecasting and achieved improved accuracy by leveraging strengths from both methodologies. Again, this is a pointer to the possibility of the strengths of integrated models in overcoming the weaknesses of univariate techniques.

Chapter 3

Methodology

3.1 Introduction

This chapter describes the data source and how it will be obtained, and the steps that will be used for fitting, forecasting and evaluating the NNAR and ARIMAX models.

3.2 Data Source and Description

The data will be obtained from CBK website `https://www.centralbank.go.ke/`.The study will also utilize temperature data from KMD.

3.3 Modelling

3.3.1 The ARMA Model

The general ARMA equation is:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{3.1}$$

Where:

- Y_t is the value of the time series at time t
- ϵ_t is the white noise error term at time t

- $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the Autoregressive terms

- $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the Moving Average terms

3.3.2 The ARIMAX Model

The ARIMAX equation can be written as:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{r} \beta_k X_{t-k} + \epsilon_t \tag{3.2}$$

Where:

- Y_t is the Dependent variable, inflation at time t.

- c is the constant term.

- ϕ_i are the coefficients of the autoregressive terms.

- θ_j are the coefficients of the moving average terms.

- β_k are the coefficients for the exogeneous variable, temperature.

- ϵ_t is the Error term at time t.

3.3.3 Steps for Model Fitting : A Box Jenkins Approach

1. Stationarity Check:

We will use the ADF test to check if inflation and temperature series are stationary.If the series are not stationary, we will apply differencing to stabilize the mean and remove trends. The differencing operation is given by:

$$\Delta Y_t = Y_t - Y_{t-1}$$

where:

- ΔY_t is the differenced series, representing the change between consecutive observations.

- Y_t is the value of the series at time t.

- Y_{t-1} is the value of the series at the previous time period $(t-1)$.

We will perform successive differencing (e.g., second-order differencing: $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$) if necessary, until stationarity is achieved. We will then use the following guidelines to interpret the ADF test results:

- **Null Hypothesis (H_0)**: The series is non-stationary.
- **Alternative Hypothesis (H_1)**: The series is stationary.

We will reject H_0 if the p-value is below a pre-determined significance level, 0.05, which will indicate that the series is stationary.

2.Identification ARIMA Parameters (p, d, q):

After achieving stationarity, we will proceed with the next step of model identification, where we will perform differencing to determine the order of differencing d, and use the ACF and PACF plots to determine p and q.

3. Incorporate Exogenous Variables:

We will then proceed by selecting external variables $(x_{t-1}, x_{t-2}, \ldots)$ that influence y_t .

4. Fit the Model:

We will estimate the coefficients ϕ_i, θ_j, and β_k using maximum likelihood or other optimization techniques.

5. Generate Forecasts:

Once the model is fitted, we will use it to forecast inflation values. The equation for forecasting is given by:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_k x_{t-k} + \epsilon_t \tag{3.3}$$

Where:

- y_t is the dependent variable.
- c the constant term.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the Autoregressive coefficients.

- ϵ_t is error term at time t.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the Moving average coefficients.
- $x_{t-1}, x_{t-2}, \ldots, x_{t-k}$ are the Exogenous variables at time t,
- $\beta_1, \beta_2, \ldots, \beta_k$ are Coefficients of the exogenous variables.

We will plug known lagged values of y_t and future lagged values of $x_{1,t}, x_{2,t}, \ldots$ into the model to compute predictions.

3.4 The NNAR Model

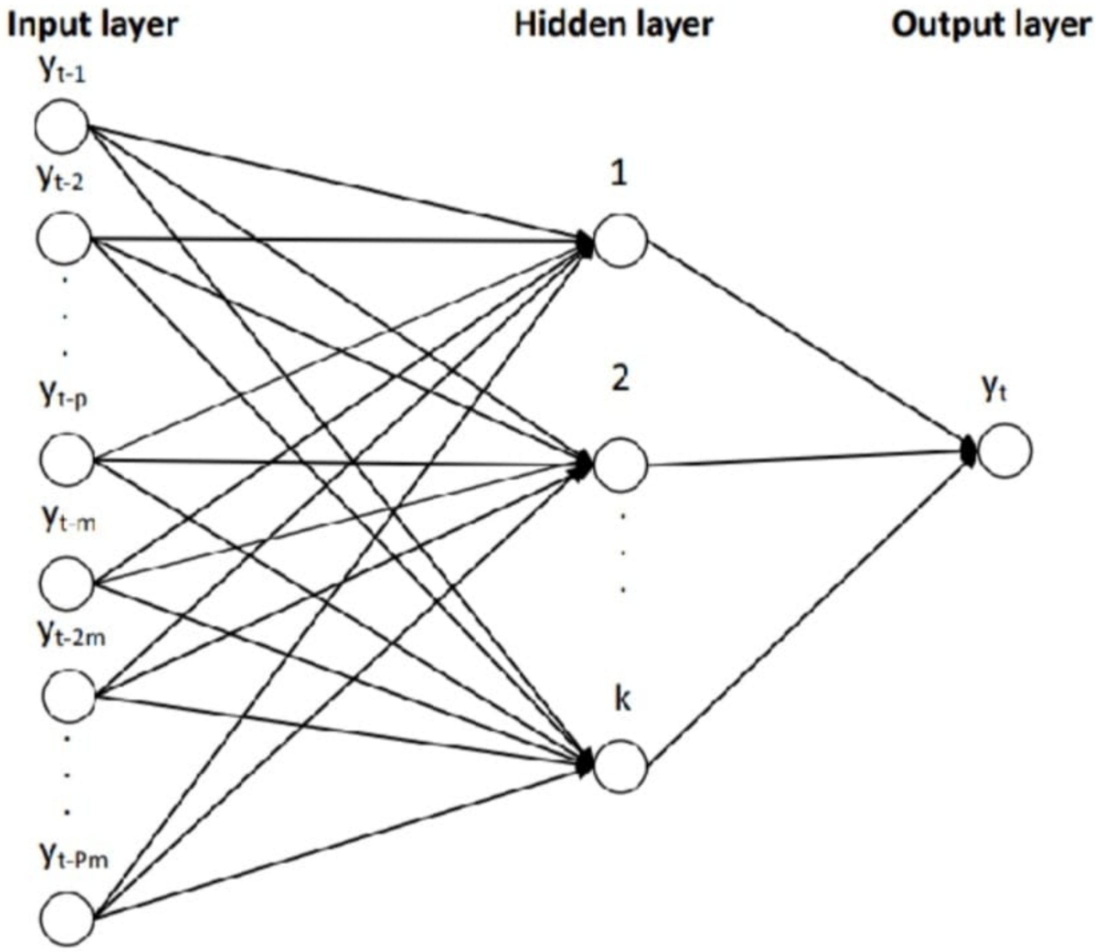

Figure 3.1: The NNAR structure

3.4.1 Model Equation

The NNAR model is represented mathematically as:

$$\hat{Y}_t = f\left(W_h \cdot \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \\ X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-k} \end{bmatrix} + b_h \right) \cdot W_o + b_o \tag{3.4}$$

Where:

- $\hat{Y}_t$ is the predicted value of the target variable at time t.
- $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ are Lagged values of the dependent variable (p lags).
- $X_{t-1}, X_{t-2}, \ldots, X_{t-k}$ are the exogenous variables (k features).
- W_h is the Weight matrix for inputs connecting to the hidden layer.
- b_h is the Bias term for the hidden layer.
- $f(x)$ is the Activation function (e.g., ReLU: $f(x) = \max(0, x)$).
- W_o is the Weight vector connecting the hidden layer to the output neuron.
- b_o is the Bias term for the output neuron.

3.4.2 Hidden Layer Computation

Each neuron's activation in the hidden layer is computed as:

$$H_j = f\left(\sum_{i=1}^{p+k} W_{h,ji} X_i + b_{h,j} \right) \tag{3.5}$$

where H_j is the activation of the j-th hidden neuron.

3.4.3 Output Layer Computation

The final prediction is given by:

$$\hat{Y}_t = \sum_{j=1}^{n_h} W_{o,j} H_j + b_o$$

where n_h is the number of neurons in the hidden layer.

3.5 Simplified NNAR Notation

The NNAR model can also be expressed in the notation:

$$\text{NNAR}(p, k, m)$$

- p is the number of autoregressive lags.
- k is the number of neurons in the hidden layer.
- m are seasonal lags (optional).

For example:

- $\text{NNAR}(12, 6)$ will be a model with 12 lags and 6 hidden neurons.
- $\text{NNAR}(12, 6, 1)$will be a seasonal model with a seasonal lag of 1.

3.5.1 Forecasting with the NNAR Model

3.5.1.1 Formulating the NNAR Model

An NNAR model can be expressed as:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}; \theta) + \epsilon_t \tag{3.6}$$

Where:

- y_t is the time series value at time t.
- $p$is the number of lagged observations (autoregressive terms).

- f is a neural network function trained on the lagged observations, parameterized by θ.
- ϵ_t is the error term, assumed to be normally distributed.

The NNAR model uses a feedforward neural network with:

- p input nodes (corresponding to lagged observations),
- At least one hidden layer with k nodes,
- A single output node that predicts y_t.

The model is commonly denoted as NNAR(p, k), where:

- p represents the number of lagged inputs,
- k indicates the number of hidden nodes in the neural network.

3.5.1.2 Steps for Forecasting with NNAR Models

1.Data Preprocessing

We will transform the time series by standardization or normalization,to ensure efficient handling by the neural network and apply log transformation if variance stabilization is required. we will then split the dataset into training and testing sets for model validation.

2.Determining Model Parameters

- *Number of Lags (p)*: We will choose p using the PACF or based on information criteria such as Akaike Information Criterion or the Bayesian Information Criterion.
- *Hidden Nodes (k)*: Typically, $k = \lceil (p + 1)/2 \rceil$, but the complexity of the data may require adjustments.

3.Train the Neural Network

We will then train the neural network using a backpropagation algorithm to minimize the prediction error.

4.Generate Forecasts

We will then use the trained NNAR model to forecast future values by iteratively feeding predictions as inputs for multi-step forecasts.

3.6 Evaluation of the Models

We will use the following evaluation metrics to test the validity of the ARIMAX and NNAR models:

- **MAE**:
$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t| \tag{3.7}$$

- **MSE**:
$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \tag{3.8}$$

- **RMSE**:
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2} \tag{3.9}$$

These metrics will help you assess how well the models will perform in capturing inflation when the climate variable is incorporated.

References

Bolanle, A. D. M. and OLUWADARE, A. (2017). Arima and arimax stochastic models for fertility in nigeria. *International Journal of Mathematics and Computer Application Research*, 7(5):1–20.

Boneva, L. and Ferrucci, G. (2022). Inflation and climate change: The role of climate variables in inflation forecasting and macro modeling.

Durevall, D. and Ndungu, S. N. (2001). A dynamic model of inflation for kenya 1974-1996. *Journal of African Economies*, 10(1):92–125.

Durevall, D. and Sjo, B. (2012). The dynamics of inflation in ethiopia and kenya. Volume Number.

Elshewey, A. M., Shams, M. Y., Elhady, A. M., Shohieb, S. M., Abdelhamid, A. A., Ibrahim, A., and Tarek, Z. (2022). A novel wd-sarimax model for temperature forecasting using daily delhi climate dataset.

Farid, M., Keen, M., Papaioannou, M., et al. (2016). After paris: Fiscal, macroeconomic and financial implications of climate change. Discussion Note SDN/16/01, International Monetary Fund.

Gathingi, V. W. (2014). *Modeling Inflation in Kenya Using ARIMA and VAR Models*. Doctoral dissertation, University of Nairobi.

John, D. L., Binnewies, S., and Stantic, B. (2024). Cryptocurrency price prediction algorithms: A survey and future directions. *Forecasting*, 6(3):1–35.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3):e0194889.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1):35–62.

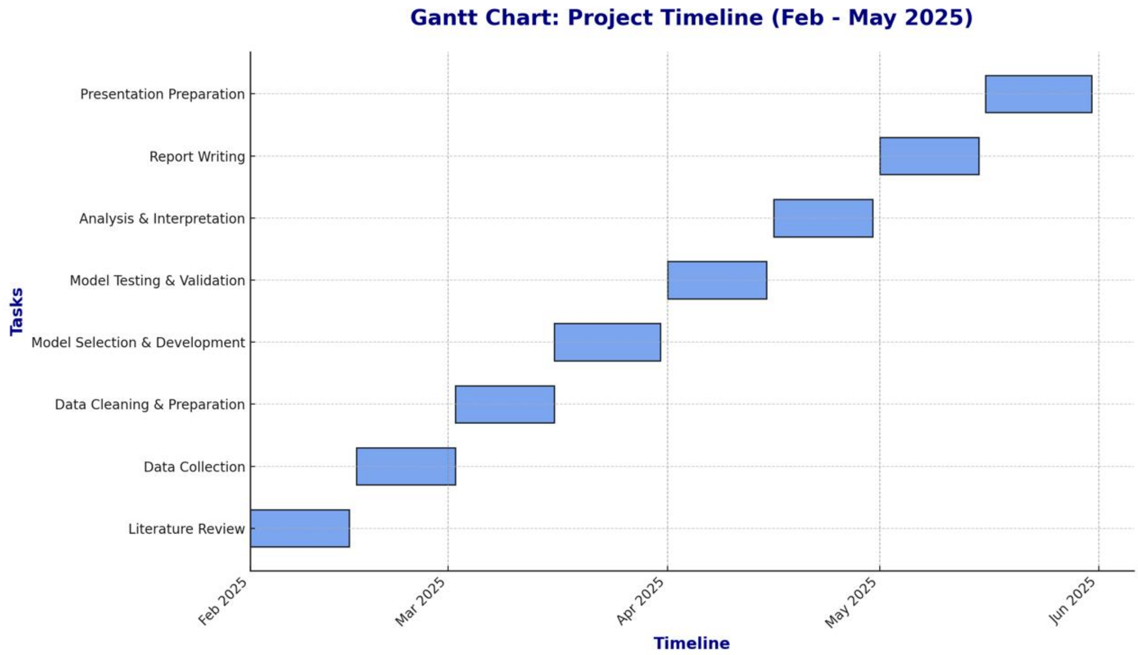

Figure 3.2: The Gantt chart

CATEGORY	DESCRIPTION	ESTIMATED COST (USD)
Data Acquisition	Subscription to financial and climate data	$50
Software and tools	Statistical software like MATLAB, SPSS, and Stata	$100
Literature review	Access to paid journals, papers and books.	$30
Printing and Materials	Printing the reports, slides and other materials	$20
Miscellaneous costs	Internet, communication and transportation expenses	$50
TOTAL		$250

Figure 3.3: The Budget